

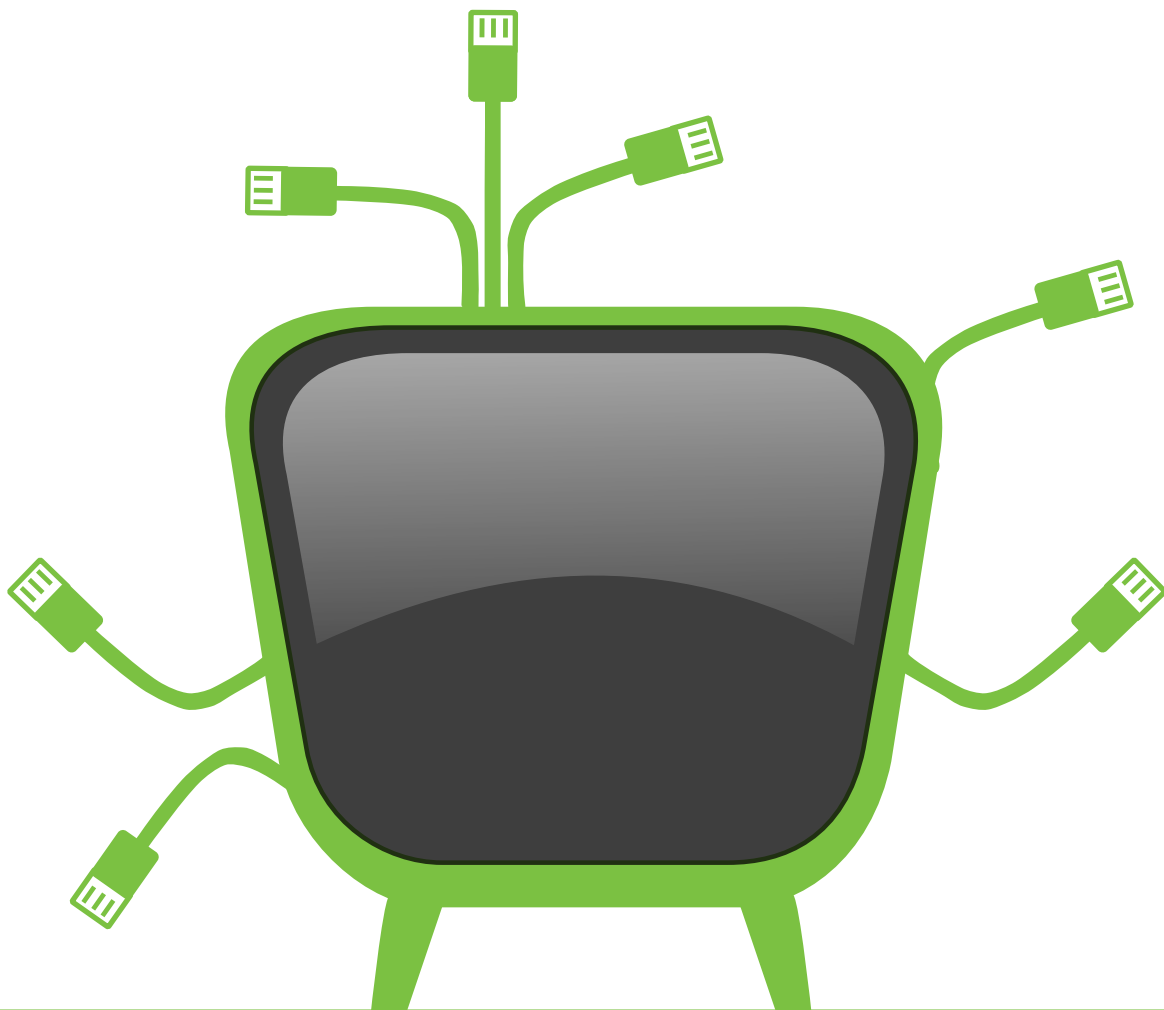


Verónica Donoso, David Geerts, Pablo Cesar, Dirk de Grooff

Networked Television

Adjunct proceedings of EuroITV 2009

Leuven, Belgium, 2009



Adjunct Proceedings Design by Verónica Donoso and Lieve Laporte
Cover Design by Dries De Roeck (c) 2009

ISBN: 9789081428002
EAN: 9789081428002

TABLE OF CONTENTS

<i>FOREWORD</i>	<i>vi</i>
-----------------------	-----------

<i>EuroITV2009 Sponsors</i>	<i>viii</i>
-----------------------------------	-------------

POSTERS

An Empirical Study of Audience-Driven Interactive Live Television on the Internet Yoshia Saito, Yuko Murayama.....	2
--	---

“TV is just one of the screens at home.” Consumers and changing TV watching Virpi Oksman, Antti Tammela, Tiina Mäkelä.....	6
--	---

MHP Interactive Applications: Combining Visual and Speech User Interaction Modes Vanessa Lobato, Gloria López, Víctor M. Peláez.....	10
--	----

Augmented TV: Experiencing Augmented Virtual TV in a Game-based PC Environment Marianna Obrist, Judith Igelsböck, Florian Förster, Manfred Tscheligi, Petra Sturm, Jochen Kranzer, Jörg Hofstätter.....	14
--	----

The embedded systems Swiss knife Luis Miguel Maia Pinto.....	18
--	----

NoTube – Making TV a Medium for Personalized Interaction Lora Aroyo, Annelies Kaptein, Davide Palmisano, Alex Conconi, Lyndon Nixon, Luca Vignaroli, Stefan Dietze, Christoph Nufer, Milena Yankova.....	22
--	----

Modela-TV: Service Personalization and Business Model Management for Mobile TV Francisco Fraile, Ismael de Fez, Juan Carlos Guerri.....	27
---	----

An Optimized Framework for Interactive TV Services Broadcasting José Antonio Hurtado-Dueñas, Bernardo Ruiz-Villalobos, J. R. Salinas, Rocío Padilla-Conejo.....	31
--	----

Non Conditional Smart Card Technology to Deliver Personalized iTV Services Giorgio Rascioni, Susanna Spinsante, Ennio Gambi.....	35
--	----

PersonalTVware: A Proposal of Architecture to Support the Context-aware Personalized Recommendation of TV Programs Fábio Santos da Silva, Luiz Gustavo Pacola Alves, Graça Bressan.....	39
---	----

From Analogue to Digital: Television Sound moving towards New Media. Stationary Sound Applications and Mobile Sound Applications	
Ahti Korhonen.....	43
Generation, Streaming and Presentation of Electronic Program Guide	
Cyril Concolato.....	46
Automatic Recommendation of (IP)TV Program Schedules using Sequential Pattern Mining	
Shinjee Pyo, EunHui Kim, Munchurl Kim.....	50
New potentials for T-Learning and informal learning with 'mashuped Television' – reinventing of Hypermedia?	
Robert Strzebkowski, Björn Münchau.....	54

TUTORIALS

'How to create an iDTV campaign' tutorial	
Koen Swings.....	59
User Experience in TV-centric Services: What to consider in the Design and Evaluation?	
Marianna Obrist, Hendrik Knoche, Damien Alliez.....	61
Ambient Media – An introduction by Case-Studies EuroITV 2009 Tutorial	
Artur Lugmayr.....	63
Designing and evaluating the Sociability of Interactive Television	
David Geerts.....	65
Standardized IPTV services in Managed IP Infrastructures: An insight on approaches from the Telco & CE industry with the focus on NGN & IMS	
Oliver Friedrich, Robert Seeliger.....	67

WORKSHOPS

Workshop 1:	
Think Positive – designing and understanding enjoyable interactive media experiences	
Think positive – designing and understanding enjoyable interactive media experiences	
Marc Hassenzahl, Susanne Sperring, Annika Wiklund-Engblom.....	71
Workshop 3:	
Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play	
Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play	
Marian F Ursu, Pablo Cesar, Doug Williams.....	75

**ANSWER: Documentation, Formal Conceptualisation and
Annotation of New Media**

Angelos Yannopoulos, Katia Savrami, Yannis Christodoulou,
Theodora Varvarigou, Stefanos Koutsoutsos, Vassilis Alexandrou.....78

**A Collective Director for Highly Interactive Viewing of Live
Sports Events**

Zhenchen Wang, Stefan Poslad, Alan Pearmain.....80

(i)TV Brings People Together?

How the Feeling of Togetherness Can Be Build

Pauliina Tuomi.....83

Good to See You -

User responses to a TV-based video telephony service

Will Seager, Hendrik Knoche.....86

**A Multimedia Content Semantics Extraction Framework for
Enhanced Social Interaction**

Giuseppe Passino, Tomas Piatrik, Ioannis Patras, Ebroul Izquierdo.....89

**Experience, narrative and interaction in TV-centred
communication**

Frank Nack.....92

Strengthen the TV Experience beyond the TV

Sara Kepplinger, Florian Förster, Judith Igelsböck, Marianna Obrist.....96

**Sharing Enriched Interactive TV Experiences with the
iNEM4U Software Framework**

Cristian Hesselman, Keir Shepherd.....99

From Social TV to Structured Communication Formats

Gunnar Harboe, Santosh Basapur.....101

Presence and mediated interaction: A means to an end?

Lizzy Bleumers, Tim Van Lier, An Jacobs.....104

Workshop 4:

Defining the Architecture for Next Generation Inclusive Television

**Defining the Architecture for Next Generation Inclusive
Television**

Richard N. Griffiths, Mark V. Springett.....108

Inclusive T-gov Application Development in Brazil

Lara Schibelsky Godoy Piccolo, Daniel Moutinho Pataca.....110

**Personalization for digitally excluded users through
t-learning**

Pablo Rebaque, Raquel Navarro-Prieto.....112

VestibaTV – An Interactive Program for Vestibular Training André Felipe Palmeira, Marcelo Fernandes de Sousa, Tatiana Aires Tavares, Ricardo Mendes C. Segundo, Guido Lemos de Souza Filho.....	114
Interactive Digital Television to enhance older people's quality of life Andrea Miotto, Jane Lessiter, Jonathan Freeman.....	116
Digital Television and the Open Mediaspace Kari-Hans Kommonen, Sanna Marttila.....	118
URC Based Architecture for Next Generation Inclusive Television Gorka Epelde.....	119
An architecture to generate automatic Brazilian sign language legends into Digital Television Systems Tiago Maritan U. de Araújo, Guido Lemos de Souza Filho, Dênio Mariz Timoteo de Souza.....	121
Accessible Digital TV: a research agenda Giorgio Brajnik.....	123
Effects of Physical Capabilities on Interaction Pradipta Biswas, Peter Robinson.....	125
Older Adults and Digital TV Amritpal S Bhachu.....	127

DEMOS

iFanz: A ubiquitous approach towards a personalized EPG Pieter Bellekens, Geert Van Kerckhove, Annelies Kaptein.....	130
DynamicTV: a new <i>Inter-tainment</i> paradigm for television Rossana Simeoni, Marina Geymonat, Roberto Antonini, Roberto Montanari, Monica Perrero.....	132
Automatic Program Extraction From TV Streams Gaël Manson, Xavier Naturel, Sid-Ahmed Berrani.....	134
Fernweh - Prototype of an IPTV-Community for Travel and Language Bjoern Muenchau, Robert Strzebkowski, Jonas Neumann.....	136
MyMedia: Dynamic Personalization of Multimedia Alexander Voß, Chris Newell, Paul Marrow.....	138
Discot: A Platform to Approach Information Society to Everyman Users Alex Ugarte, Julián Flórez, Igor G. Olaizola, David Oyarzun.....	140

Itv in industry

Designing Alphanumeric Input for iTV in a Search Context

Jay Vidyarthi, Pierre-Alexandre Lapointe.....143

Open Access IPTV Framework

Wei-Yun Yau, Fon-Lin Lai.....144

Publishing Web Content to CE Devices

Davy Van Deursen, Frank Gielen.....145

Talk to your customers through idtv

Paul Meyers.....146

User study on the impact of personalization, related content and individual preferences over interactive TV ads

Pablo Rebaque Rivas, Raquel Navarro-Prieto, Jorge Hernández Pablo.....147

Broadband Internet TV For Insurance Companies

Oscar León.....148

Accessible digital television for the visually and hearing-impaired in Flanders

Gerrit Van den Breede, Ward De Bruecker.....149

Bringing added value to iTV users?

Mirja Bächle, Kathrin Damian.....150

DOCTORAL CONSORTIUM

Ginga-NCL: Relating Imperative, Declarative and Media Objects

Marcio Ferreira Moreno.....152

Synchronization Management in DTV Applications

Romualdo Monteiro de Resende Costa.....156

Designing iTV Interfaces for Preschool Children

Ana Vitória Joly.....160

User-centered multimedia content personalization: a services composition based approach

Marcelo G. Manzato.....164

Audiovisual cultural heritage:

Bridging the gap between digital archives and its users

Guido Ongena.....168

The Next Generation of Multimedia Authoring Tools:

Telling Stories and Commenting on Media

Rodrigo Laiola Guimarães.....171

LIST OF CONTRIBUTORS.....175

Foreword

EuroITV2009, the 7th European Conference on interactive Television, reflects on one of the most radical changes in the long history of television: the shift towards the 'Networked Television', where television becomes one of the many existing interconnected devices that viewers, as users and consumers, employ in their everyday lives. It is not surprising anymore that people use their cell phones to send text messages to broadcasters, or that they use their PC to download movies that can then be watched on the big television screen. But these changes do not only imply technical challenges; being 'networked' is also impacting our everyday habits as well as our social lives. Hence, EuroITV2009 addresses important questions about the implications of shifting our focus from a conception where iTV is seen as an autonomous system independent from other networked technologies, towards a broader perspective that assumes that iTV is indeed part of a bigger set of networked devices and communities. One must, therefore, on the one hand look at this bigger context in order to meet the technical challenges that allow its evolution and on the other hand carefully take into consideration the numerous implications of such advances for 'networked' television viewers.

The reflection on the 'networked' nature of television is also present in all the 58 contributions that constitute this volume of the Adjunct Proceedings of this year's conference. These papers are grouped into six different categories including *tutorials, workshops, posters, demos, iTV in industry* and a *doctoral consortium*.

The *tutorials* section presents the description of five tutorials, covering a wide range of topics including technical aspects, design and evaluation, case studies and hands-on exercises: *"How to create an iDTV campaign"*, *"User Experience in TV-centric Services: What to consider in the Design and Evaluation?"*, *"Ambient Media – An Introduction by Case-Studies"*, *"Designing and Evaluating the Sociability of Interactive Television"* and *"Standardized IPTV services in Managed IP Infrastructures: An insight on approaches from the Telco & CE industry with a focus on NGN & IMS"*.

The *workshop* section contains the position papers of the three workshops presenting new methods, applications, design and perspectives of different aspects of interactive TV: The first workshop: *"Think Positive – designing and understanding enjoyable interactive media experiences"* focuses on an experience-oriented perspective on the current practices in the design of iTV and related cross- and interactive media. Its main aim is to collect, share and discuss design ideas and solutions to enhance users' experience. The second workshop *"Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play"* aims at exploring new forms of TV-mediated communication between groups of people (such as family and friends) separated by space and/or time that would lead to improving the maintenance and fostering of their social relationships and, ultimately, of their feeling of being together. The third and last workshop: *"Defining the Architecture for Next Generation Inclusive"* is a space to reflect on the new challenges associated with inclusivity and accessibility present in digital television. In particular, by exploring the issues and problems for inclusive design in digital television, and proposing methods, enhancements and adaptations, this workshop

will contribute to the development of new architectures for television reception that will enable much more creative solutions.

The *poster* section exhibits several papers associated with technical aspects of interactive television as well as with users' experience and consumers' perspectives. The *demo* section comprises demonstrators, prototypes, implemented services, platforms and new technologies related to interactive television. The *iTV in industry section* provides an interesting overview of existing business-oriented case studies in different parts of the world. Finally, the *doctoral consortium* paper section contains a selection of high level research work performed by young PhD researchers worldwide.

We would strongly like to encourage you to have a deeper look at the work present in this volume so as to deepen your knowledge of cutting edge research in the field of interactive television.

Finally, we would like to thank all those who made this conference possible: the authors present in this volume, the members of the organization committee, the program committee, the sponsors, and last but not least, our local organization team. Thank you all for your invaluable help, support and commitment!

Leuven, June, 2009

Verónica Donoso
EuroITV2009
Adjunct proceedings
editor
CUO, IBBT /
K.U.Leuven
Belgium

David Geerts
EuroITV2009
Program Co-Chair
CUO, IBBT /
K.U.Leuven
Belgium

Pablo Cesar
EuroITV2009
Program Co- Chair
CWI
The Netherlands

Dirk De Grooff
EuroITV2009
General Chair
CUO, IBBT /
K.U.Leuven
Belgium

EuroITV 2009

SPONSORS

INTERDISCIPLINAIR INSTITUUT VOOR BREEDBAND TECHNOLOGIE



www.ibbt.be



Philips Home Control

Philips Home Control is a global leader in remote control technology and manufactures remotes for OEMs.

Offering complete innovative solutions for all possible applications & segments, ranging from IPTV to cable & satellite, consumer electronics, home automation and pc-based applications, Philips prides itself with the broadest worldwide customer base.

Whether choosing an off-the-shelf solution or a totally original remote control that matches your product or brand identity, Philips is the right partner; building on the experience and skills of our designers, engineers and sales team.

For more information: www.remotecontrol.philips.com
remotecontrol@philips.com

PHILIPS
sense and simplicity

Turning passive viewers into active consumers?



FIND OUT HOW
AND VISIT OUR BOOTH AT



Interactive Services on Multiple TV Platforms | www.zappware.com

video-on-demand | time-shift TV | personal video recording | network-PVR | catch-up TV | start-over TV | push-VOD | music-on-demand | electronic programme guide | personalised TV | content recommendations | social networking | personalised widgets | games | interactive advertising | advertainment | ad insertion | voting | viewer polls | audience measurement | enhanced TV | digital teletext | TV site | walled garden | red button | broadcaster portal | video-on-demand | time-shift TV | personal video recording | network-PVR | catch-up TV | start-over TV | push-VOD | music-on-demand | enhanced



zappware



POSTERS

An Empirical Study of Audience-Driven Interactive Live Television on the Internet

Yoshia Saito

Iwate Prefectural University

152-52 Sugo, Takizawa, Iwate 020-0193, Japan

y-saito@iwate-pu.ac.jp

Yuko Murayama

Iwate Prefectural University, Japan

152-52 Sugo, Takizawa, Iwate 020-0193, Japan

murayama@iwate-pu.ac.jp

ABSTRACT

Traditional video services on the Internet are more in the broadcasting service nature such as streaming and video-on-demand (VoD). Recent services incorporate more of the interactive nature of network applications such as easy video sharing and those with a chat function. We have proposed an audience-driven broadcast service model in which audiences can send their wish to a broadcaster such that they like to see some specific objects while broadcasting as well as the broadcaster can reply back to the request. We implemented a prototype system for audience-driven live broadcasting and evaluated based on the results from the experimental broadcast at our university campus festival with two hundred audiences. This paper reports our experiment and findings.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces] Synchronous interaction

General Terms

Design, Experimentation

Keywords

Interactive television, Internet broadcast, Audience-driven live TV

1. INTRODUCTION

Broadband Internet services enhance video streaming on the Internet to a great extent. People without technical knowledge on broadcasting can contribute and obtain video contents easily through the web-based video-sharing services such as YouTube [1]. Recently, some broadcast services provide users with the functions to communicate with broadcasters and audiences using the interactive nature of the network; Ustream.tv [2] and Stickam [3] are typical ones. Those services provide live broadcast function and a chat function such that broadcaster and its audiences can communicate with each other while watching live streaming. Moreover, personal broadcasting is so easy. One could produce an original live program using widely available software with an economical web camera and a personal computer. We presume that the more live broadcast services we have, the more popular small and middle-scale personal broadcasting will become.

In our research, we have been trying and conducting experimental Internet broadcasting in practice [4]. Through the experiments we have found that it is quite difficult to provide audiences with the contents with quality. We are not professional

broadcaster with a huge amount of budget and technical knowledge after all. What we need is something we can afford yet we can entertain the audiences. To solve this problem, we proposed an audience-driven broadcast service model which enables audiences to request a broadcaster to execute some actions and an interactive broadcasting system, AdlivTV (Audience-Driven LIVE TV system) for the application of the model [5]. We implemented an early version of our prototype system. The prototype system, however, had only a few simple functions and was not evaluated in a practical environment.

In this paper, we report the design and implementation of our latest prototype system and the evaluation based on a practical experiment. The prototype system has several functions for audiences to make a request to a broadcaster. We evaluate the system based on our experimental broadcast at our university campus festival in October 2009 for two days. The paper is organized as follows. In the next section, we describe related work. Section 3 introduces the system model of the AdlivTV. Section 4 presents the design of the prototype system of the AdlivTV, its system architecture and user interface. Section 5 reports the experiment using the prototype system and its results. Section 6 gives some conclusions and our future work.

2. RELATED WORK

Interactive television (iTV) [6] is a research area to give interactive features to television in order to break away from a traditional one-way broadcast service model. The iTV provides audiences who watch the TV with various interactive functions with shared experiences. Typical iTV systems have chat functions such as text chat and voice chat to let audiences communicate with each other so that they have shared experiences [7, 8, 9]. They can talk with the other audiences who watch the same TV programme and feel empathy. Harboe et al. [10] proposed a Social TV system which lets audiences to know what TV programmes the other audiences are watching through ambient displays. The system makes the audiences aware of context of their friends and family. Shamma et al. [11] developed a synchronized video player for online videos to provide a shared experience. They added the video player into an instant messenger so that the IM users can watch the online videos chatting with each other. Drugge et al. [12] studied telepresence which provides an experience as if audiences were in the same place and offers various interactive functions to the audiences for the shared experience.

The AdlivTV is different from these iTV work in that it aims to provide audiences a shared experience as if they were the members of the broadcaster. The audience would be involved in broadcasting in more active manner than with telepresence,

because they could send what they would like to watch, to the broadcaster.

3. AUDIENCE-DRIVEN LIVE TV SYSTEM

The AdlivTV provides several interactive communication functions between audiences and a broadcaster to allow the audiences to send a request to the broadcaster and gives them a right to control indirectly broadcasting operations. The motivation of the AdlivTV is to let the audiences participate in the broadcast more actively and keep them interested.

Figure 1 shows the system model of the AdlivTV. In this model, we define broadcaster as a group of members who work together to broadcast a live content such as server operators and camera operators. Firstly a broadcaster starts broadcasting a live video streaming to the Internet. Secondly audiences watch the live video streaming on their computer or TV-like screen. Thirdly the audiences can request some actions to the broadcaster. For example, the requests would be “Look here”, “Move here”, “Zoom up/out” and so on. The broadcaster waits for the audiences’ requests. The audiences send their requests to the broadcaster using input devices such as a mouse, a keyboard, a microphone and a TV-like controller. The requests are represented by various ways not only chat. Fourthly when the audiences’ requests are arrived, the broadcaster should respond the requests as much as possible. The requests and responses can be received by all audiences in order to realize shared experience. While broadcasting, the request and response cycle are repeated.

The scope of the AdlivTV is small and middle-scale personal broadcasting and the target users are inexpert broadcasters neither with any budget nor professional knowledge to produce broadcasting contents. The AdlivTV provides the inexpert broadcasters with effective broadcast skills since they can have direct feedbacks from the audiences while broadcasting.

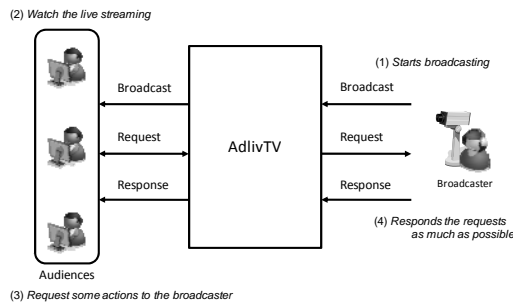


Figure 1. System Model

4. IMPLEMENTATION

We designed a prototype system of the AdlivTV to conduct a practical experiment. The implementation of the prototype system is based on Adobe Flash so that audiences can watch live video streaming on their web browser without specific software. In this section, we describe the system architecture and a user interface with interactive request functions.

4.1 System Architecture

The prototype system has a streaming server, several clients and a camera operator. Figure 2 shows the system architecture of the prototype system.

We use Red5 v0.7.0 final to build the streaming server. The Red5 is an open source Flash server written in Java. The streaming server receives video data from the camera operator and broadcasts it to clients. A server program also works on the server to handle requests from the clients and manage the system.

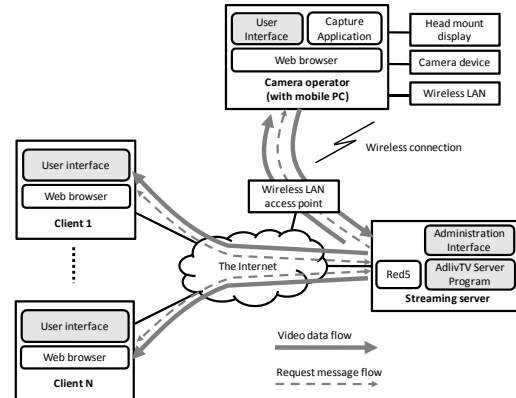


Figure 2. System Architecture

The camera operator has a Head-mounted display (HMD), a camera device and a small mobile PC equipped with wireless LAN (WLAN). The camera operator can check the current broadcasting and audiences’ request through the HMD. The mobile PC is connected with a WLAN access point and the camera operator can walk around to comply with audiences’ various requests.

When the client accesses to a web page for the broadcasting, a Flash application for the audiences is launched. At first, the client connects to the AdlivTV server program with a unique number. The server program generates an audience’s ID using the received number as a seed and sends it to the client. The client keeps the unique number in the Cookie on the web browser and can receive same ID from the server each time. The ID is used for system logs identifying audiences. The logs maintain connected time, disconnected time and audiences’ request with the IDs. After obtaining the ID, the client establishes several connections to the server and they can send requests through an original user interface for the AdlivTV watching live video. In the next subsection, we describe detail of the user interface

4.2 User Interface

Both audiences and broadcasters use the same user interface to watch a live video with the prototype system. Figure 3 shows the user interface of the AdlivTV. (Note: It is the Japanese version and the underlined English parts are for the purpose of the explanation.) The user interface provides a live video player and two functions for the audiences to make a request.

The first function is a chat. This popular communication function lets the audiences and the broadcasters exchange text messages each other.

The second function is a request icon function. Those icons are presented in the menu and the users can select one of them by a click and the request is sent to the broadcaster. We have the following icons; (1) **Look**: Focus the camera onto the specific object, (2) **Move**: Move to the specific location, (3) **What**: Ask about the specific object, (4) **Attention**: Pay attention to the specific object, (5) **Like**: Understand that the specific object is my

favorite, (6) **Talk**: Talk to the specific person, (7, 8) **Zoom in/out**: Zoom in/out at the current angle, (9, 10, 11, 12) **Look left/up/down/right**: Turn to the specific direction, (13) **Good**: Continue current good broadcasting, (14) **Bad**: Change current bad broadcasting. (Note: The *Good* and *Bad* icons increase counters shown on the bottom of the video field.) Those icons are unique with our system. The icon function provides the audiences and the broadcaster with easy and intuitive communication tool. It shows the icons selected by the audiences on the video in a few seconds. It is useful for stylized requests.

All audiences can use the icon function anytime and the icons are shown on the video field of the all audiences. Sometimes several audiences may use the request icon function at the same time and several icons are shown on the video field. In this case, the broadcaster, i.e. the camera operator judges what request should be applied and executes one of them.

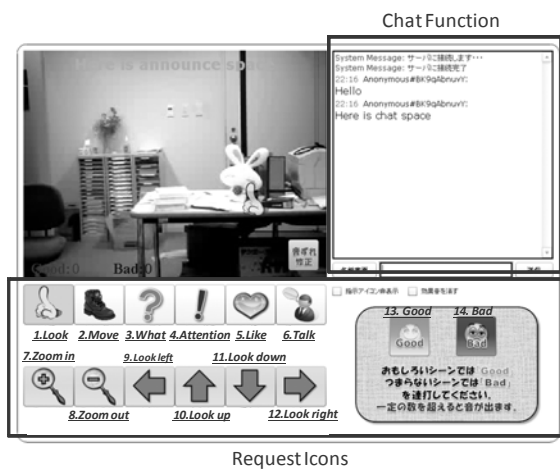


Figure 3. User Interface

5. EXPERIMENT

We carried out a practical experiment to study how effective the audience-driven broadcasting works and in what way audiences make requests to the broadcaster while finding issues of the system in the practical situation. The experiment was conducted in the campus festival of Iwate Prefectural University on October 25 and 26, 2008.

5.1 Environment

The festival was held in the campus of Iwate Prefectural University. Figure 4 shows the network configuration in the experiment. The AdlivTV server was placed in a broadcasting server room of our laboratory and connected to the Internet at 100 Mbps. Many live performances such as singing and dancing were held at the main stage. We set up a WLAN access point equipped with a router function near the stage. The WLAN access point connected to the AdlivTV server via LAN in the University. In our experiment, we used two cameras, a mobile one and a fixed one so that the broadcaster could switch the video footage when the camera operator changed the battery of the mobile PC or some problems were happened. We used mainly the mobile camera in the experiment.

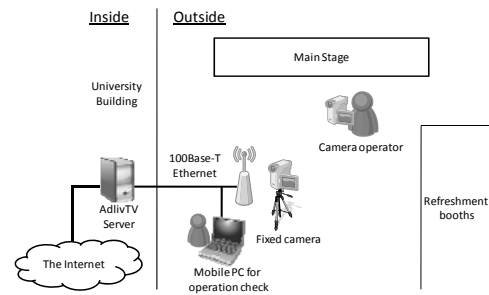


Figure 4. Network configuration

5.2 Results

At first, we counted the total audience on October 25th from 10:00 to 15:00 and 26th from 10:00 to 17:00, 2008. Note that we excluded broadcasters' access from all the results. The total number of the audience was 93 on the first day and 200 on the second day. It was small-scale broadcasting in the scope of the AdlivTV and the number of audiences was quite adequate for our experiment.

We analyzed the changes in the number of audiences for the two days to study increase and decrease of the audiences. Figure 5 shows the changes per 60 seconds. In our experiment, we had two types of cameras, viz. the mobile one and the fixed one. We needed quite often to switch the mobile one to the fixed one due to the various troubles of the mobile one such as a need to change the battery of the mobile PC attached to the mobile camera. Vertical solid and dotted lines in the graph show the time when we switched to the mobile camera and to the fixed one respectively. We can see that audience rate decreased when we switched to fixed camera and increased when mobile camera was used. From the result, we find that audiences are sensitive for the camera work and broadcasters could get more audiences if they provided their live programme with audience-driven and dynamic camera work.

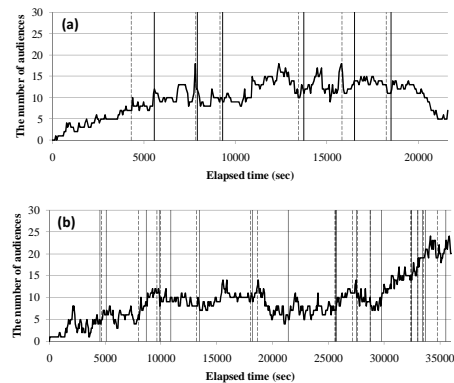


Figure 5. Change in the number of audiences per 60 seconds
(a): 25 Oct. 2008, (b): 26 Oct. 2008

We analyzed the use of request functions as well. As for the chat function, the audiences used the chat function 146 times on the first day and 63 times on the second day. The request messages included in the chat were 17 on the first day and 4 on the second day. Examples of the request message includes "Please go into the specific building", "Move to the best shooting place for fireworks" and "Too much noise, fix it please" and so on.

However, the request messages were only ten percents of the total chat messages and the rest were comments on the broadcasting. From the result, we can see that the chat function was not used frequently to send requests to the broadcaster but it was useful for complaints and some requests in the specific situation.

As for the icon function, we analyzed change in the use of the request icons per 60 seconds and total use of each request icons. Figure 6 shows the use of the request icons. From the graphs, we can see that the use of the request icons is concentrated in a short time and it happens suddenly. Since the camera operator judged which requests should be responded, the concentration of requests sometimes made the camera operator confused. We should support the camera operator to be able to select an appropriate one out of hundreds of requests.

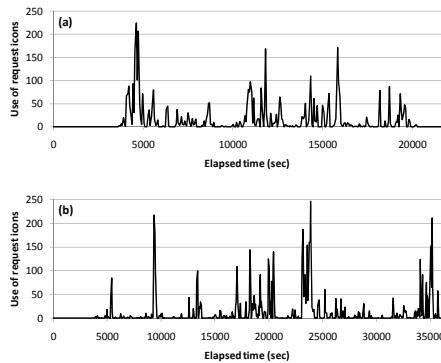


Figure 6. Change in the use of the request icons per 60 seconds (a): 25 Oct. 2008, (b): 26 Oct. 2008.

Table 1. Total use of each request icon

Type of request : Total use			
Look:1230	Like:617	Look left:1460	Good:1155
Move:602	Talk:592	Look up:190	Bad:1218
What:943	Zoom in:617	Look down:90	
Attention:466	Zoom out:555	Look right:888	

Table 1 shows the total use of each request icon. The frequently-used icons include “Look”, “What”, “Like”, “Look left”, “Look right”, “Good” and “Bad”. The Look icons would be useful for the audiences to control camera works. The “What” icon was important to acquire information from the broadcaster and the “Like”, “Good” and “Bad” icons helped to present audiences’ impression to the broadcaster. On the other hand, “Look up” and “Look down” were almost never used in the experience because the main stage was horizontally wide and it did not need vertical camera works. From the results, we found that the audiences could use the request icon function easily and they were involved actively in the broadcast using the function.

We also found an issue of the request icon function for the camera operator to respond requests. The audiences used “Talk” icon to a person on the main stage and they continued to use the icon although the camera operator could not execute the action. Since the camera operator could not reply that the request was not executable, the audiences kept sending the useless requests. It would not satisfy the audiences if their request were ignored many times. The system should provide a response function so

that the camera operator can decline expressly inexecutable requests with an easy operation.

6. CONCLUSION

We introduce our work on the AdlivTV and reported our implementation of a prototype system and its evaluation by a practical experiment at our university campus festival. From the experiment, we had four findings; 1) broadcasters could have more audiences if they provided audience-driven and dynamic camera work, 2) the icons for requests were used in a short period of time and in a burst way, 3) the audiences used the icon function easily and were actively-involved in the broadcast by the use of the function, 4) a response function should be provided for the camera operator to decline inexecutable requests.

As future work, we will study support functions for the camera operator since workload of the camera operator is intensive with the current prototype system. We need to provide more support functions for the camera operators so that they can select an appropriate one out of hundreds of requests and decline requests expressly with easy operation.

7. REFERENCES

- [1] YouTube, <http://www.youtube.com/>.
- [2] Ustream.tv, <http://www.ustream.tv/>
- [3] Stickam, <http://www.stickam.com/>
- [4] N. Takamizawa and Y. Murayama, “Use of Telepresence in Informal Broadcasting over the Internet”, In Proc. of the ACM SIGMM 2004 Workshop on Effective Telepresence, pp. 12-15, 2004.
- [5] Y. Saito, Y. Isogai and Y. Murayama, “AdlivTV: An Interactive Broadcasting System for Audience-Driven Live TV on the Internet”, In Adjunct Proc. of EuroITV2008, pp. 217-221, 2008.
- [6] J. Jensen, “Interactive Television – A Brief Media History”, In Proc. of EuroITV2008, pp. 1-10, 2008.
- [7] D. Geerts, “Comparing Voice Chat and Text Chat in a Communication Tool for Interactive Television”, In Proc. of NordiCHI '06, pp. 461-464, 2006.
- [8] T. Coppens, L. Trappeniers, M. Godon, “AmigoTV: towards a social TV experience”, In Proc. of EuroITV2004.
- [9] H. Weisz, S. Kiesler, H. Zhang, Y. Ren, R. Kraut and J. Konstan, “Watching Together: Integrating Text Chat with Video”, In Proc. of CHI'07, pp. 878-886, 2007.
- [10] C. Harboe, C. Metcalf, F. Bentley, J. Tullio, N. Massey and G. Romano, “Ambient Social TV: Drawing People into a Shared Experience”, In Proc. of CHI'08, pp. 1-10, 2008.
- [11] D. Shamma, M. Bastea-Forte, N. Joubert and Y. Liu, “Enhancing online personal connections through the synchronized sharing of online video”, CHI'08 extended abstracts, pp. 2931-2936, 2008.
- [12] M. Drugge, M. Nilsson, R. Parviainen and Peter Parnes, “Experiences of Using Wearable Computers for Ambient Telepresence and Remote Interaction”, In Proc. of SIGMM 2004 Workshop on Effective Telepresence, pp. 2-11, 2004.

“TV is just one of the screens at home.” Consumers and changing TV watching

Virpi Oksman

VTT Technology Centre of Finland
P.O. Box 1300
33101 Tampere, Finland
+358505883070
virpi.oksman@vtt.fi

Antti Tammela

VTT Technology Centre of Finland
P.O. Box 1300
33101 Tampere, Finland
antti.tammela@vtt.fi

Tiina Mäkelä

VTT Technology Centre of Finland
P.O. Box 1300
33101 Tampere, Finland
tiina.makela@vtt.fi

ABSTRACT

In this paper, we present findings of empirical studies on insights of end users' attitudes and hopes towards peer-to-peer (P2P) networked television. The paper aims to look at the changing trends of consumers' media usage and television watching.

General Terms

Human Factors, Experimentation, Security

Keywords

Peer-to-Peer TV, User Experiences, Internet TV, Social Media

1. INTRODUCTION

Consumers are using multiple technologies such as their PCs, media players and mobile phones to download and watch news and other TV contents at home and on the go. The growth of peer to peer (P2P) networks and social web has increased the possibilities of content sharing. Moreover, the traditional TV set has been developed towards a networked social medium which can be used at the same time chatting with friends and surfing on the Internet. In addition, several other kinds of interactive TV services are being developed, which move TV away from a linear broadcast model.

Thus in many respects, TV is moving from a collective medium to a more personal and active approach used with different kind of ICT devices both at home and in various situations outside the home. There is evidence that among other things people value mobility, personality, diversity and real-time effects as important properties of media services. [5]

Clearly, television is changing and diversifying its forms. The activity of viewing programs is converging with peer-to-peer and other interactive activities such as chatting, shopping, voting and playing. [3]

As a result, we can see a growing trend of media usage beyond the traditional TV watching. A shared TV experience is still important; however, there is an increase in the use of different screens and multiple TV sets at home. More often, each household member is using individual mediums and different kinds of interactive media services.

Moreover, new forms of media creation, delivery, and consumption have been evolving. The possibilities to create and consume a variety of contents and services are increasing. In addition, there are new forms of community and sociability being developed, for instance the possibility of commenting on programs with friends outside your home.

Also platforms for community-supported media annotations and remixes have been implemented. Remixes are made over a range of genres, with over a third showing thematic unity and a quarter showing some attempt at narrative. [7]

Furthermore, the issues of security are gaining new significance with the new on-line services as people want to protect their information privacy.

Nevertheless, it is not quite clear how certain P2P-services are experienced as a part of future TV and how people would like to use them in general. The real significance on social and interactive TV depends on understanding of the uses, purposes and problems of such systems. [1, 9] Therefore, this paper will present and discuss the findings of empirical studies on insights of end users' attitudes and hopes towards the television of the future.

2. RESEARCH QUESTIONS

To find out what kind of qualities and functions users would prefer on the television of the future, this paper focuses on three research themes based on empirical research:

1. *Content*: What kind of media contents would users want to consume, and how much would they be willing to pay from these services?
2. *Sense of community*: What forms of sociality and community would people like to engage in over the P2P services? Would users like to chat with their friends while watching television?
3. *Security*: What are the things users are worried when using this kind of services?

In the project we have adopted a human-centred design approach in which end-users are engaged at an early stage of development. Based on this knowledge it is possible to constitute a user's requirement specification for future P2P services.

3. METHODS

3.1 Scenario Survey

Responding to the research questions outlined above requires the gathering of diverse research material. Our aim was to gather data from same areas of interests with different research methods; with scenario survey and focus groups. This makes possible to compare the results with different research methods. Even though the sample sizes are quite small it is possible to make some interesting findings on user's attitudes to P2P services for the future development.

Scenarios are personalized fictive stories which make the usage of devices more concrete to users. This method was used in the survey because P2P-services are not familiar to all users. Therefore, scenarios help users to understand what is meant with different interactive qualities and functions.

Scenarios are descriptions of a single user who is using a device or a service in order to achieve certain result in certain circumstances and in certain time period. They are simplified descriptions without technical details and they are written from the user's point of view. Moreover, scenarios can also be used to describe the usage of the devices of the future. They are best suited for early stages especially for the requirement specification. [4, 2, 6]

The scenario survey consisted of visualizations of different use situations and a questionnaire concerning the qualities or the functions mentioned in the scenario. 65 users responded to the survey questionnaire. The respondent's age range was from 15 to 63 years. Their mean age was 31 and 33 of respondents were women and 32 were men. The scenario survey was conducted in VTT's open web lab called Owela.

3.2 Focus Groups

Focus group is a good method for finding out attributes, opinions and attitudes. Technique is especially applicable in finding out first impressions when the service is new and when the use population is not specified. Focus group is suitable in the early developing stage for determining the user requirements. It is recommendable to have more than one focus group because the outcome of single focus group may not be representative. [8, 4]

Three focus groups were held in order to collect users' opinions and attitudes towards P2P-services. The first focus group consisted of users with technical experience who were all users of social media like Facebook, MySpace and YouTube. The second focus group consisted of middle age and older users and the third group were young adults.



Picture 1. A scenario presenting the possibility for watching together and commenting on programs with P2P-services.

4. FINDINGS

The findings from the scenario survey and focus groups supported each other. The main message of the focus group was that the service must be user friendly in order to succeed. In addition, users appreciated freedom of choice, functionality, reliability and picture quality. They also noted that there should not be too much spam. Users thought that there may be advertisements but not too many and they should not disturb the watching experience.

According to our expectations, the users were not willing to pay much from the service. Unwillingness to pay was explained with the current television fee in Finland. It would possible to pay 1-2 euros for sport events and maximum 5 euros for a movie. They were mostly interested in paying videos with single payment with credit card or mobile phone. More options for selecting and watching videos were wished for. Users suggested also that it should be all the time possible to cancel your order. You should be billed only from the period that you've viewed. This way users could more easily and carefree order TV-programs.

4.1 Contents

The main feedback was that the users wanted to order programs that are not currently available from regular television channels for example movies, documents and series. They also wanted to have the opportunity to order classics and to watch series premieres.

Moreover, the users were interested in ordering language and educational material and specific things that interested them for example videos regarding their hobbies. Users also pointed out the importance of up-to-date news.

According to survey many users were interested in surfing on the internet while watching TV (Figure1), men more often than women. However, in the focus groups some of the users were in the opinion that they just want to watch television and not to do anything else while watching. Others thought it was a good idea because they are recently doing the same thing with different devices. Especially younger users already chatted with the laptop computers and watched television at the same time. Some users said that they are already using the Internet and the television together because they have a digital television receiver in their computer. For some users, the activity of viewing TV programs is converging with other on-line activities: *“Television is just one of the screens at home. It could be handy to for example manage your banking account on television”*, said 21-year old man.

To read emails on television might have at least novelty value for younger users. There were also suspects, that it might addict television viewers so that you are all the time examining new mails. Users noted that it should be possible to disconnect this kind of television services so that it doesn't disturb during a thrilling movie. *“It could be nice if it doesn't suspend your viewing. It might spare some time”*, said 22-year old woman.

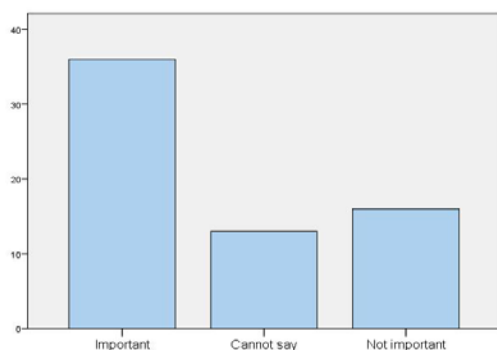


Figure1. How important is it for you to watch TV and surf on the Internet at the same time?

4.2 Sense of community

The survey results and the focus groups showed some signs of changes in television watching. About one third of the scenario survey respondents considered that it is still important to view television together with your friends or with your family at home. A little bit more said that it is not important anymore (Figure 2). An explanation for that might be that quite many respondents thought that they want to share the viewing experiences with their friends who live elsewhere. Results from the focus groups were similar. Attendees thought that it is quite common to discuss TV programs afterwards in the workplaces during coffee breaks. Some of the focus group members used also text messages to discuss TV during the viewing. TV is still strongly a social medium, but the social community might be elsewhere than at home, on the couch.

The younger users reflected that their parents are interested in what is happening in their lives. Parents would like to get videos from their children's life and whereabouts. Not surprisingly younger users were not willing to send videos to their parents. However, the younger group was ready to send videos to their friends. The older users wanted to share videos with their relatives and family. Especially if relatives were living abroad shared videos would be a good way to keep in touch. Older users also thought that it would be much more convenient to watch videos on television than on a computer screen. According to the users sharing videos from local events could be useful for different hobby communities. Users were also interested in watching live music from local happenings.

Especially the younger users thought that P2P services are best suited for people who do not own a computer or it could be a channel to communication for lonely people. However, they also commented that if the service would be easy-to-use enough, it would be suitable for all kinds of people.

The users were shown a scenario presenting the possibility for watching together and commenting on programs with P2P-services with your own community of friends (see Picture 1). The younger group said that this kind of chatting and commenting suits especially well to sports. In fact the younger user group was used to view and comment television programs on a computer screen at the same time. They were watching for instance Formula races on the left part of the screen and discussing it on the right part. Also current affairs programs were mentioned as a possible chatting environment.

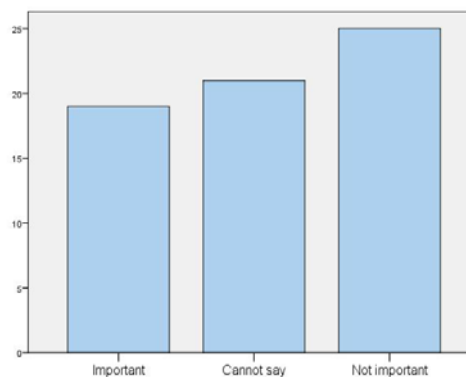


Figure 2. How important is it for you to watch television together with your family or friends?

4.3 Security

Especially the older users were concerned about safety issues related to P2P services. They were worried that their private and personal messages and information would spread to public. Because TV is a medium with a mass audience, they insisted that they have to be absolutely sure who can see your messages.

Users thought it would be good if they could put their own videos under a password, but there were still privacy concerns. Clearly,

personal information privacy is one of the most important issues of P2P television services.

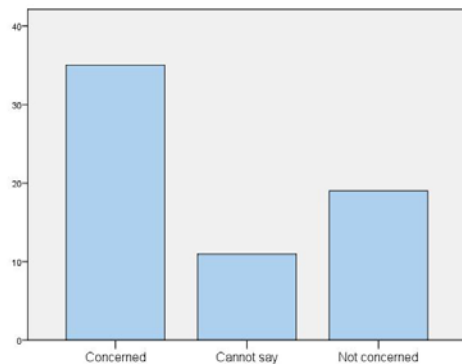


Figure 3. Would you be worried with this kind of service that the videos would spread to wrong hands?

5. CONCLUSION

All the focus groups discussed the changing concept of television. Television and computers seem to integrate in future homes. Should we actually just talk about screens in different places?

7. REFERENCES

- [1] Harboe, M., Harboe, G., Massey, N., Metcalf, C., Wheatley, D. & Romano, G. 2007 Perceptions of value: The uses of social television. In Cesar, P. Chronianopoulos, K., Jensen, J. F. (Eds.) *Interactive TV: A shared experience*. /Heidelberg: Springer, 125-116.
- [2] Helander, M. G., Landauer, T. K. and Prabhu, P. V. 1997. *Handbook of human-computer interaction*, North-Holland.
- [3] Livingstone, S. 2004. The challenge of changing audiences. Or, what is the audience researcher to do in the age of the internet? *European Journal of Communication*. Vol 19(1), 75-86.
- [4] Nielsen, J. 1993. *Usability engineering*. AP Professional.
- [5] Oksman, V., Ollikainen, V., Noppari, E., Herrero, C., Tammela, A. 2008 'Podracing': experimenting with mobile

One may find it quite surprising that quite many people do not consider watching TV together with their family or friends as an important activity anymore. It appears that the traditional TV related sociability is collapsing as each family member would rather choose to watch their favourite programs with their own time and with their personal media devices and services. More likely, while watching TV, people keep several mediated interactions which do not require the parties to be in the same space at the same time [9]. Yet, this hardly means that people would not like to participate in social interaction at all. Probably there are some other hobbies, activities and common interests which are taking the role of the traditional TV watching for family.

Obviously, consumers' media usage, behaviour and experience is changing if each family member needs media of their own and more likely the number of screens at home will increase. Yet, it is not quite clear what extent and what kinds of media technologies users are willing to accept as a part of their home environments. There is still lack of recent studies of home contexts and users' behaviours [10]. In the future, we will continue the research on home environments and gather comparable research material from different countries.

6. ACKNOWLEDGMENTS

Part of this work is supported by European Union P2P Next project. We would like to thank all our partners for their co-operation

TV content consumption and delivery methods. *Multimedia Systems*. Vol 14, Number 2/July. Springer

- [6] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. 1994. *Human-computer interaction*, Addison-Wesley.
- [7] Shaw, R. & Schmit, P. 2006. Community annotation and remix: a research platform and pilot deployment. 89-98, ACM, New York, NY, USA, Santa Barbara, California, USA.
- [8] USINACTS – Usability in the Information Society CD. 2000. The USINACTS project was funded by the European Commission under the ACTS programme.
- [9] Hjarvardt, S. 2008. The mediatization of society. A theory of the media as agents of social and cultural change. *Nordicom Review* 29(2), 105-134.
- [10] Obrist, M., Bernhaupt, M. and Tcheligi, M. 2008. Interactive TV for the Home: An ethnographic study on user's requirements and experiences. *Int. Journal of Human-Computer Interaction*, 24(2), 174-196.

MHP Interactive Applications: Combining Visual and Speech User Interaction Modes

Vanessa Lobato

Fundación CTIC, R&D Department
Parque Científico Tecnológico Gijón
33203, Asturias, Spain
+34 984291212

vanesa.lobato@fundacionctic.org

Gloria López

Fundación CTIC, R&D Department
Parque Científico Tecnológico Gijón
33203, Asturias, Spain
+34 984291212

gloria.lopez@fundacionctic.org

Víctor M. Peláez

Fundación CTIC, R&D Department
Parque Científico Tecnológico Gijón
33203, Asturias, Spain
+34 984291212

victor.pelaez@fundacionctic.org

ABSTRACT

This paper proposes an architecture for the development of interactive digital television systems accessible via voice. The proposed architecture is applicable to any interactive digital television application developed under the MHP standard, which runs on an interactive receiver equipped with return channel via Ethernet. The user needs only a mobile device with audio I/O, capable of running a speech synthesis and recognition system which follows the W3C standards SSML and SRGS respectively. This work includes the formalization of the architecture to meet this objective and validates it through two use cases in the scopes of eHealth and Digital homes.

Categories and Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces - Graphical user interfaces (GUI), Input devices and strategies, Interaction styles, Voice I/O.

General Terms

Design, Human Factors

Keywords

DVB-T, Multimedia Home Platform (MHP), interactive digital television (iDTV), speech interaction, speech recognition (ASR), speech synthesis (TTS)

1. INTRODUCTION

Digital Terrestrial Television (DTT) has generated great interest related to accessibility of the audiovisual media because of multiple possibilities of DTT in terms of services and applications. Multimodal interfaces, customization and/or standardization [1] are key factors in the search for a universal design of interactive systems. This paper takes this approach, focusing on the use of standards as well as on multimodality, combining voice and graphic modes. It is important to highlight that voice interaction is an indispensable requirement for users with visual impairments, as well as one of the most natural interaction modes for the general public [2].

In 2005, the Spanish Technical Forum of Digital Television published a list of requirements for digital television accessibility [3], which includes the requirements for interactive systems and voice-based interfaces. Nevertheless, there are difficulties in

putting these requirements into practice due to a lack of standardization and hardware constraints.

Standardization is a key element to boost iDTV services as evidenced by the Multimedia Home Platform¹ (MHP) initiative of the Digital Video Broadcasting (DVB) consortium, widely accepted in Europe.

Hardware limitations are primarily due to the fact that receivers (or set-top boxes, STB) tend to have the minimum hardware features to meet their specific purposes. These features are generally insufficient for voice software requirements. Moreover, STBs rarely have an audio input, which is an essential requirement in automatic speech recognition (ASR). Aware of these limitations, in 2003 the *National Centre for Accessible Media* (NCAM) published a guide to create spoken menus for STBs and DVDs², destined to both software developers and hardware manufactures. This guide includes information not only about hardware constraints, but also about the advantages of using Text-To-Speech (TTS) technology compared to pre-recorded voices in terms of flexibility and cost.

There are several publications that address the inclusion of a voice interface in DTT applications [4] [5] but these are usually limited to the scope of electronic program guide (EPG) navigation. There are also some studies about the quality of the ASR in iTV [6], which propose different strategies for error recovery in the ASR. Some studies into accessibility are based on improving the graphical user interface, making it more suitable for people with visual impairments [7], while others deal with the humanization of the navigation experience [8] (e.g. using as remote control a mobile device). Further relevant initiatives are accessible Emplea-T [9] and IntegraTV-all [10]. Both of these address the inclusion of voice interaction in an application that has been operating in a real environment. The first one includes speech output to an iTV application, although it uses pre-recorded messages instead of TTS technology; the second one uses ASR and TTS on proprietary technology and specific hardware.

All the works described above have a design of specific prototypes for particular applications, depending in many cases on specific hardware such as a Media Centre or a specific STB.

¹ <http://www.mhp.org/>

² <http://ncam.wgbh.org/resources/talkingmenus/>

The search for a global architecture should first address the hardware limitations discussed above. One option is to manufacture a custom-built digital TV receiver, which would cover the processing needs of the system. Another alternative is to use an additional hardware device with audio I/O, connected to the digital receiver and capable of running the voice software. Because of the first option is a time-consuming approach to obtain a specific product which would probably not be compatible with the products of other manufacturers, the work presented here is based on the second alternative. This approach has advantages in terms of costs and interoperability.

The main objective to achieve in this work is the inclusion of voice interaction in interactive applications build with MHP. The resulting applications will allow people with visual disabilities to interact with MHP applications in their own homes.

2. SYSTEM ARCHITECTURE

This paper proposes a distributed architecture with two basic blocks (see Figure 1): the MHP interactive application, running on the STB; and the voice control module, running in an external and portable device with audio I/O (e.g. a PDA). Both functional blocks are connected through a network interface. Speech recognition and synthesis systems will be implemented on the mobile device, using its own audio inputs and outputs. This implies that the communication between the voice control module and the interactive application will not be used to transmit audio signals. Instead, the text for the synthesizer, the grammars that govern the process of recognition or the recognized text are the data to be transferred.

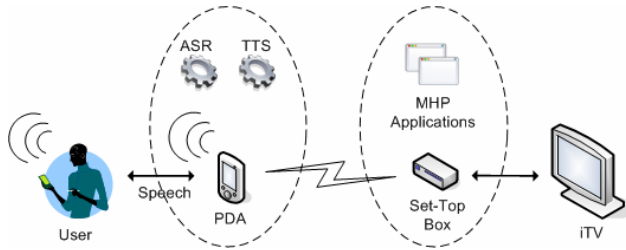


Figure 1. Global system architecture.

One of the main advantages of the proposed architecture is derived from the distribution of voice processing in a separate hardware device (e.g. a PDA). This feature makes it possible to combine a fully reusable voice module with any application that has network interface, regardless of the application scope.

Thus, three main steps have to be taken into account to design a MHP application with voice interface and based on the approach described above:

1. Development of the interactive digital television application using the MHP standard
2. Connection of the interactive application with the voice control module
3. Integration of the voice interface with the graphical interface, usual in this kind of applications, and synchronization of the two interaction modes.

2.1 Multimodal Interactive Applications

Once the architecture that specifies the hardware elements involved and their responsibilities have been defined, it is necessary to establish a mechanism to integrate and synchronize multiple modes of interaction within the interactive application for DTT. The interactive MHP applications considered here, interaction with the user based on voice and visual elements, can be defined as multimodal applications according to the classic definition of a multimodal interface: *“that system which processes two or more combined input modes”* [11].

Besides the application logic and the modules for the different interaction modes, the overall architectures considered in multimodal systems [11] [12] provide a number of common modules. These modules are the multimodal integrator or fusion module, the dialogue manager and the response planner or fission module. Because the expected interaction mechanisms of the system presented here are considered as alternative and individual elements, we propose a union of the fusion module, the dialogue manager and the fission module in a single module that will assume the three tasks. This simplifies the necessary architecture and makes a more efficient and suitable deployment for a MHP environment.

In addition, the proposed architecture for the interactive applications isolates the user interface from the logic necessary to integrate and sync the interaction modes (see Figure 2). The architectural pattern called Model View Presenter (MVP) has been used, as its use provides independence between the interaction modes (*views*), and also perfect synchronization through the *presenters*.

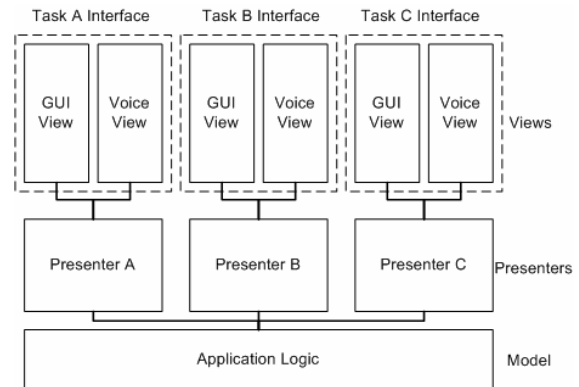


Figure 2. Interactive application architecture.

The *graphical views* (GUI) are the different interface screens, with the usual graphic components in this kind of applications, such as buttons, lists, texts, etc. The *voice views* contain the grammars for speech recognition with the commands that the user can give to the system and the messages to be synthesized. Both, graphical and speech views, translate the relevant events (button activations, reception of voice commands, etc.) into uniform actions that the presenter will trigger depending on the origin and the event type (e.g. dispatching response events to other views).

Following this pattern, events and user actions taken in different views will be transferred to the presenter, which will assure consistency and synchronization between all of the views. In this way, the presenter is fully responsible for the fusion, fission and dialogue management modules of a classic multimodal system.

2.2 Speech Control Module

In the proposed approach, speech technology lies in a common PDA which communicates with the interactive digital television application. Nevertheless, it is important to note that any hardware device with audio I/O, such a mobile phone or a tablet PC, can also be a valid option.

Speech technology has been selected with support for the standards JSGF³/SRGS⁴ and SSML⁵, to specify ASR grammars and TTS respectively. For this reason, it is possible to use any ASR-TTS engine that supports these standards.

The speech control module is based on a client-server architecture, where the client is the DTT interactive application and the server resides in the PDA. Through the exchange of information, the speech technology resident in the PDA is controlled from the STB according to the application state. This control includes updating of the active grammar in the voice recognition module (ASR), the definition of the text to synthesize and other options such as language selection in multi-language applications. On the server side, the PDA executes the operations of audio input/output associated with the ASR and TTS systems, based on information previously received from the STB. The voice commands which are identified by the ASR are sent instantly from the PDA to the STB, where they are interpreted.

Voice control on the PDA has been designed according to the method Push to Talk or 'press and talk,' because of its robustness against noise in the ASR process.

The interactive application development should be independent from the implementation details necessary to establish the communication with the synthesis and recognition services. To this end a component for MHP that encapsulates all the implementation and provides a simple interface with the necessary operations has been developed. This component is responsible for managing communications, starting the service in the PDA or the remote device, and translating the operations into the necessary network commands.

3. EXAMPLES OF USE CASES

To validate the architecture and design proposed we have taken two examples of interactive applications for DTT which verify the two basic requirements defined; that is, to be based on the MHP specification and to have an architecture based on the Model-View-Presenter pattern. The first example is an application for monitoring, recording and analyzing sleep quality in the home, developed in the scope of an European project⁶ (Figure 3). The second example focuses on the monitoring and control of digital homes (see Figure 4). In both cases, the speech interface not only

recognises the actions indicated by the user through voice commands, but also gives audio feedback to the user about ongoing actions (for example, the user might hear “*your data is being sent to your health centre*”).

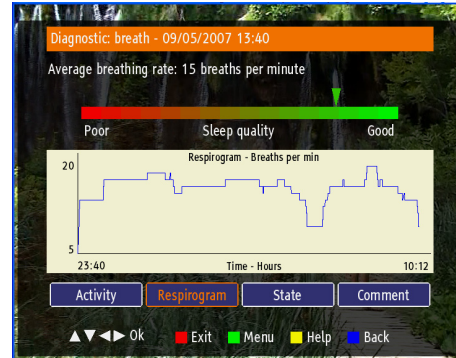


Figure 3. Application 1 (“Diagnostic results” screen).



Figure 4. Application 2 (“Digital elements” screen).

In both applications, the user can set the language in which the different elements and texts are visualized by pressing a key on the remote control unit. In addition, that functionality is available in voice-based user interaction. In this way, it is possible to change the language in which commands are recognized and texts are synthesized on-line, regardless of the speaker (user) or the characteristics of each type of voice (male/female). The languages used in the testing phase were Spanish and English.

Another useful feature in the ASR is to use grammars within semantic content, as in the SISR standard. In other words, the ASR module can return directly recognized voice commands (*Start*, *Stop*, etc.) or a semantic representation previously associated with those commands. Using grammars with a semantic content simplifies the processing of the information received from the ASR module in the interactive application. When using a grammar with multiple choices activated simultaneously in a multi-language environment (English, Spanish, etc.) there are two main options that can be considered:

1. Without semantic representation, the message from the PDA to the STB is the whole recognized chain (for example, ‘*Set Spanish language*’). In this case it is necessary to process the message in the interactive application, to first establish the type of command and then its associated value, if this exists.

³ <http://www.w3.org/TR/jsgf/>

⁴ <http://www.w3.org/TR/speech-grammar/>

⁵ <http://www.w3.org/TR/speech-synthesis/>

⁶ <http://www.estiic.org> – SMITH Subproject

2. With semantic representation, each command is assigned a code (e.g. 1 to *Start*, 2 to *Set language*, etc.). Thus, when a voice command is recognized, the PDA sends a message consisting of a pair '*code [value]*' to the STB.

This second approach based on semantic information has been applied successfully in the developed prototypes. The initial tests carried out have demonstrated that the user can easily perform all the available functionalities by voice commands, receiving feedback via synthesized voice messages.

The results of the implementation evaluation show that the speech recognition accuracy is good for a low noise environment due to the use of the built-in PDA microphone which has short range sensitivity. Using a push-to-talk strategy to launch the speech recognition process simplifies the user interaction with the system because the user only has to touch the PDA screen to begin the process. Two limitations were found to affect the implementation of the system. Firstly, the recurrent use of the help command in order to remember the available voice commands (this is not a dialogue based system). Secondly, the complex interactions necessary to control some graphical widgets like selection lists. The latter limitation indicates that although the proposed architecture allows voice interaction to be included in existing applications, in order to improve the user experience it would be advisable to redesign some of the graphical user interfaces, or at least to consider different interaction schemes for the voice mode. This topic will be addressed in future work.

4. CONCLUSIONS AND FINAL REMARKS

In this work an architecture for the development of interactive digital television systems with speech interface, according to the MHP standard is proposed. This architecture allows voice interaction to be included in any interactive application that follows an architectural pattern similar to the Model View-Presenter and runs on a digital television receiver equipped with a return channel via Ethernet.

The hardware alternative proposed on this paper presents several advantages in comparison to others based, for instance, on the purchase of a STB built ad-hoc. In this sense, the approach presented here allows the use of any common STB which is MHP compliant to run the interactive application, and any external device with I/O audio to execute the TTS and ASR modules.

There is no doubt about the multiple advantages of multi-modal interface from the user experience point of view. In this sense, it is important to highlight the total synchronization between the different interaction modes available in the interactive applications used to validate the proposed design, both input (graphic/touch and voice), and output (graphic and voice). This synchronization means that any user interaction with the system (visual or voice based) implies an update of all the views whenever necessary.

Future work includes work on the multimodal interaction, combining all the modes simultaneously, validating the proposed architecture with more use cases and involving more users during the testing phase in order to identify new functionalities and requirements. This future work will contribute to improved versatility and potential of the global system and, of course, to a better user experience.

5. REFERENCES

- [1] Gill, J. M. and Perea, S. A. 2003. Accessible Universal Design of Interactive Digital Television. In Proceedings of the 1st European Conference on Digital Television (Brighton, UK, April 2-4, 2003). EuroITV'03.
- [2] Neto, J. and Farinazzo, V. 2008. Non-Functional Requirements to Voice User Interface on Interactive Television: an Initial Study. Proceedings of the International Conference Interfaces and Human Computer Interaction (Amsterdam, The Netherlands, July 25-27, 2008). IADIS'08.
- [3] Foro Técnico de la TV Digital. Accesibilidad en Televisión Digital para personas con discapacidad. Octubre, 2005.
- [4] Wittenburg, K., Lanning, T., Schwenke, D., Shubin, H. and Vetro, A. 2006. The Prospects for Unrestricted Speech Input for TV Content Search. In Proceedings of the working conference on Advanced Visual Interfaces (Venezia, Italy, May 23-26, 2006). AVI'06. ACM, pp. 352-359.
- [5] Ibrahim, A. and Johansson, P. Multimodal Dialogue Systems for Interactive TV Applications. In Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (Pittsburgh (PA), U.S.A., October 7-10, 2002). pp. 117-222.
- [6] Berglund, A. and Qvarfordt, P. 2003. Error Resolution Strategies for Interactive Television Speech Interfaces. In Proceedings of the IFIP TC13 International Conference on Human-Computer (Zurich, Switzerland, September 1-5, 2003). INTERACT'03. IOS Press, pp. 105-112.
- [7] Rice, M. and Fels, D. 2004. Low Vision and the Visual Interface for Interactive Television. In Proceedings of the 2nd European Conference on Interactive Television (Brighton, UK, 31 March – 2 April, 2004). EuroITV'04.
- [8] Cereijo, A., Sala, R., Ahmad, S. and Rahman, M. 2005. Beyond the remote control: Going the extra mile to enhance iTV access via Mobile devices & humanizing navigation experience for those with special needs. . In Proceedings of the 3rd European Conference on Interactive Television (Aalborg, Denmark, 30 March – 1 April, 2005). EuroITV'05.
- [9] Martín, C.B., Merchán, J.M., Jiménez, D., Menéndez, J.M. and Cisneros, G. 2007. Accesibilidad a la Televisión Digital Interactiva. In Proceedings of the 2º Congreso de Accesibilidad a los medios audiovisuales para personas con discapacidad (Granada, Spain, June 21-22, 2007). AMADIS'07. Real Patronato sobre Discapacidad, pp. 67-77.
- [10] Ceccaroni, L., Martínez, P., Hernández, J.Z. and Verdaguer, X. 2005. IntegraTV-4all: an Interactive television for all. In Proceedings of the 1st Symposium on Ubiquitous Computing and Ambient Intelligence (Granada, Spain, September 14-16, 2005). UCAmI'05. In J. Bravo, J. Alamán and T. Riesgo (Ed.).
- [11] Oviatt, S. 2002. Multimodal interfaces. Handbook of Human-computer Interaction, J. Jacko y A. Sears (Ed.), Lawrence Erlbaum, New Jersey.
- [12] Trung, H. 2006. Multimodal Dialogue Management – State of the art. Human Media Interaction Department, University of Twente.

Augmented TV: Experiencing Augmented Virtual TV in a Game-based PC Environment

Marianna Obrist, Judith Igelsböck, Florian Förster, Manfred Tscheligi
ICT&S Center, University of Salzburg
Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{marianna.obrist, judith.igelsboeck, florian.foerster2, manfred.tscheligi}@sbg.ac.at

Petra Sturm, Jochen Kranzer, Jörg Hofstätter
ovos realtime 3D GmbH
Waaggasse 2a / EG
1040 Vienna, Austria
{ps, jk, jh}@ovos.at

ABSTRACT

This paper presents the results of a first user evaluation of the Augmented Virtual TV prototype designed to augment the TV program experience. It gives the user the possibility to apply the information learned in a traditional TV program in a PC based application using a game format, which is closely connected to the topics featured in the broadcasted TV program. In this paper, we present the work carried out putting a special focus on user experience and user acceptance by using a combination of evaluation methods: a workshop with integrated focus group based on defined tasks, free exploration, two validated questionnaires and group discussion.

Categories and Subject Descriptors

H.1.2 Models and Principles: User/Machine Systems—Human factors; H.5.1 Information Interfaces and Presentation: augmented and virtual realities; H.5.2 User interfaces: Theory and methods.

General Terms

Design, Human Factors.

Keywords

Augmented TV, game-based PC environment, user evaluation, user experience, user acceptance.

1. INTRODUCTION

The experience of watching TV is no longer only connected to a TV set, but expanded into new broadcast technologies like PCs, mobile phones or handhelds. New possibilities and devices to consume TV content implicate the risk of a poorer user experience [1]. They further enable a switch from passive to active viewing habits [2], slashing dependency on broadcast schedules. The intertainment paradigm, for instance, suggests that interfaces should allow users traditional ways of consumption and at the same content access known from new media environments [3].

Within this paper we present the concept of Augmented Virtual TV (short Augmented TV). It contributes to a future understanding of TV program navigation systems and interactive TV format design. Aware of the initial cross media scope it offers interesting input for the content production community perspective as well as for e-learning. A special interest of Augmented TV is to enrich the understanding of the potential of virtual architecture freed from accustomed but nonessential spatial arrangements. Thus, Augmented TV goes beyond traditional

interactive TV (iTV) concepts, and crosses platforms by introducing the TV content into a virtual augmented world, currently realized in a game-based PC environment. Within a first user study the concept of Augmented TV was evaluated addressing the following questions: First, to what extent do the users consider the idea of combining a TV program and a Computer Game as meaningful? Second, how is the Augmented TV prototype perceived in terms of User Acceptance and User Experience? Additionally some user requirements regarding means of communication and community were collected.

2. RELATED WORK

Augmented TV connects elements of broadcast television, virtual environments and gaming.

2.1 Augmenting the TV Experience

Looking at the history of iTV, different characteristics like video on demand, offering additional content and enhancing the overall TV experience (e.g. e-commerce, text chat or internet surfing) can be detected [4, 5]. Various experiments on introducing iTV applications have been conducted. However, there is still a lack of research how the users' TV experience could be positively supported with interactive services and applications.

For developing a successful iTV program a high semantic connection is of importance [6]. The producer should strive for a close connection between the TV content and the interactive application, augmenting the experience of the viewers but leaving them within the cognitive horizon of the program. In our research we aimed to evaluate if the users experience of Augmented TV was closely connected to the TV show. Augmented TV exceeds traditional interactive contributions to a TV program by offering the viewers a complete virtual world they can immerse.

2.2 TV and Virtual Worlds

The idea to share TV content in virtual worlds has recently received a boost as media and ICT companies are expanding their presence into virtual worlds. Virtual world applications are most of the time connected to existing brands (e.g. MTV's Virtual Laguna Beach¹ or Nickelodeons Nicktropolis²). They feature avatars for users, shopping, games and strong social networking components. Other examples include the outlets of media companies in second life. The Australian Broadcast Corporation

¹ http://www.mtv.com/ontv/dyn/laguna_beach/vlb.jhtml

² <http://www.nick.com/nicktropolis/game/index.jhtml>

offers an ABC Island³, where ABCs' audio and video content in a social environment as well as watch live screenings and concerts can be consumed. Despite these practical examples for TV related virtual world applications, there is no data on popularity, users' experiences, reception or acceptance of these offered applications available.

Initial research on user participation in TV programs via shared virtual worlds was conducted by Benford et al. [7] in the late 90ies. Until 2001 four programs were developed that gave the viewers the possibility to take part in a virtual world via the Internet. The aim was to produce TV shows of the events in the virtual environments and broadcasted them live. The concluded that established design principles can facilitate the coherence between virtual world and the broadcast: namely simplicity in concept, clear roles, cooperation between different participants and interaction through proximity.

The principles of entertainment in inhabited TV were recently discussed by Fanciulli [8]. He integrates knowledge taken from massively multiplayer online game experiences. He stresses the importance of emotional engagement and virtual sensation through the avatar that is essential for entertaining audiences in the environment of virtual spaces.

2.3 Gaming Environment

Apart from the virtual world concept, Augmented TV is also based on gaming concepts, which support learning experiences. To optimize the game experience educational games should combine a story that integrates the challenges into a larger context [9] with internally consistence and fairness [10] as well as balance between attractive elements and educational objects in order to optimize the possibility of players experiencing flow [11].

An example of mixing TV and a game was developed by BBC: Bamzooki⁴ is a mixed reality TV show which features a toolkit for kids. The pilot went on air on CBBC in March 2004. A sequel was produced in 2008. The downloadable toolkit allows kids to set up their own spider robot avatars. Some of them are selected for the show where they compete against each other. In the TV studio the virtual spiders fight on a table surrounded by their creators. The main point is, that the children are unaware of this augmented reality contrary to the TV viewers.

3. AUGMENTED TV

Augmented TV is developed as part of a national funded project. The system is mainly designed for content formats, which derive advantages from heavy user interaction like game shows, reality TV, science formats and others TV programs. It enables an active role of the user and the personification of TV contents. Augmented TV is driven by the concept of having additional benefits (like learning) by playing. The user can personalize the system and explore it through playing. Thus, Augmented TV unites TV with gaming and Internet using the advantages of each part. To realize the architectural concept of virtual space, the real time 3D engine Quest3D was chosen. Quest3D offers high rendering performance in combination with the flexibility to feed the special needs of this format. This whole range of possibilities is necessary in order to implement the inhomogeneous future designs and contents within the overall structure of the project.

³ <http://www.abc.net.au/services/secondlife/default.htm>

⁴ <http://www.bbc.co.uk/cbbc/bamzooki/>

Although an implementation of Augmented TV on the TV screen via a set top box is aspired, technical requirements will not allow this experience for the next years. Augmented TV consists of a client application, which can be installed on every PC suitable for common games. All clients communicate with each other and a single server application enables a virtual multi-user environment. After starting the client, the players select and modify their avatars, which are visualized as small robots. When entering the virtual world (see Figure 1) the user meets other players and can access topic related games based on the TV show.



Figure 1: Virtual world of Augmented TV

According to the subject of the TV show, portals open up within the virtual world, which lead the players to game worlds. The interaction with the content of the TV show takes place in these game worlds. Two case examples of TV shows were accessible for and realized in the Augmented TV prototype, namely mountain climbing and wind parks. In the first scenario the players are asked to find the best route up a mountain (see Figure 2) by setting marks. Afterwards the scored result can be measured by an ideal route (to enable a learning effect) and compared with other players' results. In this case the avatar does not play an active role and does not enter the game world (he stays in the "normal" virtual world). In the second example the player can learn to build an efficient wind park to gain as much energy as possible while taking other factors like economic efficiency or population satisfaction into account (factors were discussed in the accompanying TV program). In this large scale game world the avatar shows up because there is real need to navigate through.



Figure 2: Realized Case Example 1: "Mountain Climbing".

4. STUDY SET UP

In the following the methodological set up for the evaluation of the recent Augmented TV prototype and the results are presented.

4.1 Used Method & Participants

For gaining first insights into the perception of the Augmented TV idea a two hours workshop with integrated focus group was conducted (see Figure 3). The workshop was organized with six participants, three male and three female, between 19 and 25 years old, each of them with specific media knowledge deriving from academic education (Multimedia Art and Technology).



Figure 3: Workshop Set Up and Participants.

The workshop consisted of four phases combining qualitative and quantitative methods. At the beginning, an explorative brainstorming phase aimed at finding out how participants imagine a hybrid between a TV program and a Computer Game without knowing Augmented TV. Subsequently the participants got to know Augmented TV in a walkthrough phase (performed in pairs), where the participants had to fill in feedback cards and small questionnaires related to pre-defined tasks (i.e. task 1: Start Augmented TV, login and customize the avatar, task 2: Enter the forum and navigate through, task 3: Play the mountain game, task 4: Free exploration, task 5: Leave the game). In addition, the feedback cards provided feedback on what the participants liked and disliked, as well as how difficult the tasks were perceived.

In the next phase, the participants were asked to fill in two standardized questionnaires, namely the AttraktDiff [12] and the Game Experience Questionnaire GEQ [13] to measure the game experience. Within the AttraktDiff questionnaire participants were asked to indicate their perception of the Augmented TV prototype by the help of antithetic word pairs (semantic differential, e.g. motivating – discouraging). All items had to be rated on a scale ranging from the negative (-3) to the positive (+3) word pole. The AttraktDiff questionnaire provides an overall impression on user experience by measuring the pragmatic and hedonic quality of a computing system. Out of 28 items four cumulative scores are built, namely for the Pragmatic Quality (PQ), Hedonic Quality – Stimulation (HQ-S), Hedonic Quality – Identity (HQ-I), and Attractiveness (ATT) (see details in [12]). Moreover, the GEQ questionnaire is consisting of 33-items addressing the following components of user experience: Immersion, Flow, Competence, Tension, and Challenge.

The workshop ended with a focus group discussion addressing the user acceptance of Augmented TV: the ease of use, the perceived usefulness and the novelty of the Augmented TV concept

4.2 Results

4.2.1 User Experience Results

The participants had most problems with the usability of the Augmented TV prototype, in particular with the control of the mountain and the windmill game. However, they liked the idea and the learning effect of Augmented TV as well as the character design. The average rating of the difficulty requested within the feedback cards was 3.02 (from 1 very difficult to 5 very easy). Task 2 “enter the forum and navigate the avatar” was rated as easiest, while task 3: “play the mountain game and afterwards return to the forum” was perceived as most difficult (average ratings per task were: task 1: 3.33, task 2: 4, task 3: 1.75, task 4: no rating (free exploration), task 5: 3). The results of the AttraktDiff questionnaire (see Figure 4) show that Augmented TV is perceived as novel and inviting (best average scores) on the one hand, as technical and isolating (lowest average scores) on the other hand.

In addition, Figure 5 shows the mean values of the four cumulative factors (PQ, HQ-S, HQ-I, ATT) in the AttraktDiff. The worst average factor score reaches Pragmatic Quality indicating that Augmented TV is not estimated as easy to control. The factor Attractiveness gets best results, which shows that Augmented TV is perceived as likeable and inviting. Nevertheless participants are not completely satisfied with the current aesthetics and suggest improvements (see 3.3.2 Acceptance Results). Hedonic Quality scores are around zero, which indicates that participants are still undecided about the value of Augmented TV.

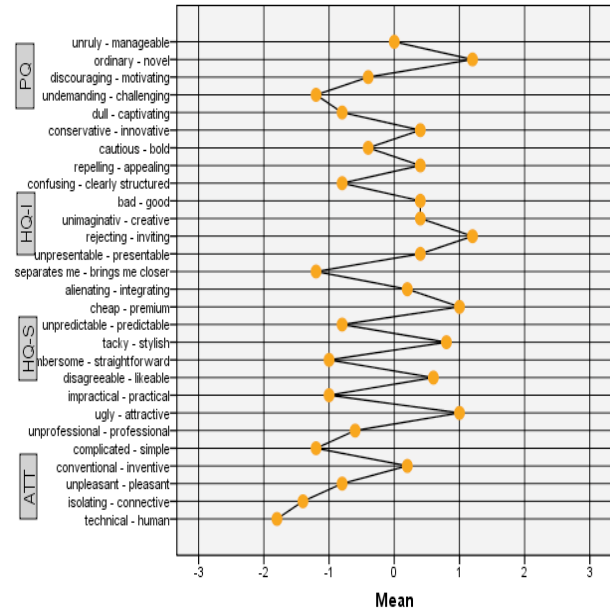


Figure 4: AttraktDiff: Semantic Profile for Augmented TV.

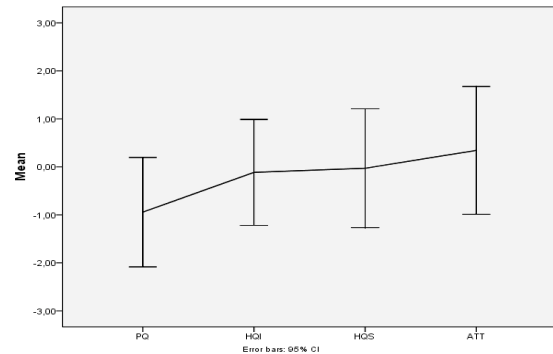


Figure 5: Average AttraktDiff Factor Scores.

The GEQ questionnaire shows that participants were not stressed or frustrated by Augmented TV at all (factor tension – see Figure 6). They perceived it as rather challenging indicating that the level of difficulty is well chosen within Augmented TV. Factor flow scores worst, which can maybe traced back to the walkthrough situation that did not allow participants getting deeply into the whole potential of Augmented TV. The factor competence, which describes the ease of use, is perceived differently, partly as okay, partly as poor. The low score of immersion indicates that the storyline around Augmented TV needs to be improved

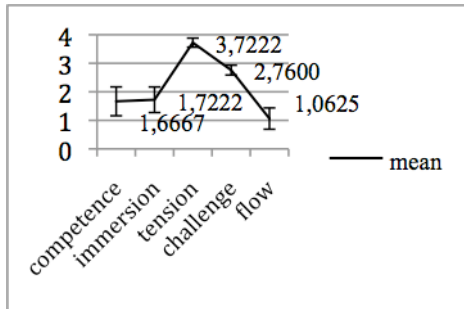


Figure 6: GEQ: Average Experience Factor Scores.

4.2.2 User Acceptance Results

User acceptance was addressed in the group discussion at the end. It revealed that participants are keen on the learning effect provided by Augmented TV. They also embrace the aesthetics, even if they found some inconsistencies in the presence of the avatar in the games and some discordance in viewing perspective. Moreover, the results of the GEQ regarding perceived ease of use (competence) were verified in the discussion: partly participants stated “controls are okay as they are now”, partly participants were not able to handle the controls and demanded more instructions. Even if they accepted the overall idea of Augmented TV, participants commented that they are missing a narrative scenario round the whole Augmented TV, explaining the reasons of why to create an avatar and why the avatars are robots.

4.2.3 Additional User Requirements

Augmented TV should offer several means of communication. Even though participants prefer chat, they like to be able to choose between several communication possibilities. Profile pages for each user are preferred as long as Augmented TV does not turn into another community application like Facebook. They stated that the last few years they have been oversaturated of community sites and thus tired of them. Finally, the participants suggested to improve the connection between a TV program and a computer game through the introduction of different game levels bounded to the TV show.

5. CONCLUSIONS & FUTURE WORK

In this paper a first user study for the recent prototype of Augmented TV was conducted. It revealed that the participants consider it as interesting, innovative and challenging, but there is room for improvement. In particular the lack of semantic connection [6] between the TV program and the Augmented TV application needs to be addressed in further iterations. The narrative embedding of Augmented TV needs to be defined and explained thoroughly to the users. At the same time an integration of elements of the TV show can support the cohesion of the TV program and the game-based PC application. In general, the participants liked the idea of learning by using Augmented TV and can imagine using it in future. Based on the results gathered due to this first evaluation, a new version is developed and will further be evaluated in a laboratory setting, mainly addressing usability issues as well as the overall user experience. The revised

prototype of Augmented TV is currently presented to selected persons in the media business, industry and education facilities in Europe and USA. The feedback will also be incorporated into future developments and will help to define next steps.

ACKNOWLEDGMENTS

We acknowledge the support of the people participating in the conducted user evaluation study.

6. REFERENCES

- [1] Cooper, W. 2008. The Interactive Television User Experience So Far. In Proceeding of the 1st international conference on Designing interactive user experiences for TV and video. UXTV '08. ACM New York, 133 – 142.
- [2] Ursu, M. F., Thomas, M., Kegel, I., Williams, D., Tuomola, M., Lindstedt, I., Wright, T., Leuridijk, A., Zsombori, V., Sussner, J., Myrestam, U., and Hall, N. 2008. Interactive TV narratives: Opportunities, progress, and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.* 4, 4 (Oct. 2008), 1-39.
- [3] Simeoni, R., Geymonat, M., Guercio, E., Perrero, M., Rapp, A., Tesauri, F., and Montanari, R. 2008. Where Have You Ended Up Today? Dynamic TV and the Inter-tainment Paradigm. In *Proceedings of the EuroiTV08*, Springer Berlin/ Heidelberg, 238-247.
- [4] Jensen, F. J. 2008. Interactive Television - A Brief Media History. In *Proceedings of the EuroiTV08*, Springer Berlin/ Heidelberg, 1-10.
- [5] Jensen, F. J. Interactive television: new genres, new format, new content. *Proceedings of the second Australasian conference on Interactive entertainment. IE2005*. Creativity & Cognition Studios Press, Sydney, Australia, 89 – 96.
- [6] Goldenberg, S. 2008. Creating Augmented and Immersive Television Experiences Using a Semantic Framework. In *Proc. UXTV '08*. ACM, 45 – 48.
- [7] Benford, S., Greenhalgh, C., Rodden, T., and Pycock, J. 2001. Collaborative virtual environments. *Commun. ACM* 44, 7 (Jul. 2001), 79-85.
- [8] Fanciulli, M. 2008. Principles of entertainment in inhabited television. In *Proceedings of the Working Conference on Advanced Visual interfaces (AVI '08)* ACM, 5-12.
- [9] Seagram, R., Amory, A. 2004. Designing Effective Stories for Educational Games. In L. Cantoni & C. McLoughlin (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004*. Chesapeake, VA: AACE, 162-167.
- [10] Rollings, A., Adams, E. A. 2003. *Rollings and Ernest Adams on game design*. New Riders, USA.
- [11] Kiili, K. 2005. Digital game-based learning: Towards an experiential gaming model. In *Internet and Higher Education*, Elsevier 8, 13-24.
- [12] Hassenzahl, M. 2005. The thing and I: understanding the relationship between user and product. In *Funology: from usability to enjoyment table of contents*, Kluwer Academic Publishers Norwell, MA, USA, 31-42.
- [13] IJsselstein, W.A., de Kort, Y.A.W. & Poels, K. (in preparation). The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games.

The embedded systems Swiss knife

Luis Miguel Maia Pinto
Novabase Digital TV
Av. do forte nº6
2790-072 Carnaxide, Portugal
(+351) 969545234
luismiguelmaiapinto@gmail.com

ABSTRACT

Current corporate environment are somehow not open to information and functionalities presentation to other channels than the World Wide Web. For ISP's this is quite common were the main distribution channel is the world wide web, forgetting most popular channels such as TV and the emerging web enabled mobile devices. This document aims to present the currently project undergoing on ZON Multimédia, the major Portuguese cable operator, where the self-care solution is being prepared to be used in web either by PC, or any embedded device (phone, pmp's, stb's, etc...). This work aims to design a simple, seamless process to allow easy deployment of web application from the pc environment to embedded devices, and also support the evangelization process of the company decision maker's by the architect, by so building a truly Swiss knife for the system architects.

Categories and Subject Descriptors

D.2.2. [Design Tools and Techniques]: Programmer workbench and Features – *system analysis, information integration, system design.*

General Terms

Algorithms, Design, Human Factors.

Keywords

Algorithms, Design, Human Factors, Embedded Systems, System Analysis, Information Integration

1. INTRODUCTION

Traditionally corporate environment has a global difficulty of envisioning innovation that may sprout from the system designers, or even from the teams closest to the clients. During a long partnership with ZON Multimédia, I was able to identify the ZON desired to allow their clients to access the existing customer care infra-structure using embedded devices. (TV, video-game consoles, mobile devices, etc ...)The first problem encountered was the fact that corporate decision makers didn't have the technologically awareness of embedded devices to have the right perception of what it should be done, in order to pursue the objectives they had in mind. In order to attend this desired objectives of ZON Multimédia it was needed to devise a working strategy that would allow the proper evangelization of the decision maker, on the subject of designing solution for embedded devices, and also a solution on what process should be followed to design and implement a platform that supported multiple distribution channels of several functionalities

2. GLOBAL STRATEGY

In order to solve this problem the adopted strategy was:

1. System analysis: Gain knowledge of the overall system to be supported in the embedded systems, and identify the targeted embedded systems.
2. Usage modeling: Gain knowledge from the clients that uses the system identifying what are the major functionalities used by them, and what functionalities they expect to be available in a embedded system.
3. Design interface strategies for the identified embedded systems
4. Design construction processes
5. Test strategies
6. Usage Evaluation

Step 1 aim's to evaluate the system to be ported on to the targeted embedded systems, mapping all the functionalities that are available to users. Step 2 evaluates the overall system usage, identifying what are the most used functionalities by the clients. These results will allow identification of the most likely candidate functionalities to be available in the embedded systems. Considering that on embedded environments it can be easily identified a direct relation between overall system functionalities and global system acceptance, looks that adopting the most used functionalities will be a good strategy. Another fact must be considered when designing embedded systems, it states that the simpler the interface is the easier is the user acceptance of the application (this fact simply states that if an interface is simple, the less cognitive effort it requires, then is much easier to the user to learn the interface mechanics). Summing this we will have a relation that tells us the bigger the usage of existing functionalities that are transported for the embedded system, and the bigger simplicity in which those are presented, most likely the system is going to have wide acceptance on the target users.

Step 3 presents the strategies that application and interface designer's need to follow when building interfaces for embedded systems.

On step 4 the development pipeline is the target of the investigation, on how the development process should be managed.

Step 5 and step 6 describes how the platform is tested prior to live launch and post launch evaluation on platform usage.

3. SYSTEM ANALYSIS

After some initial analysis of the overall ZON Multimédia self-care platform the following architecture was identified:

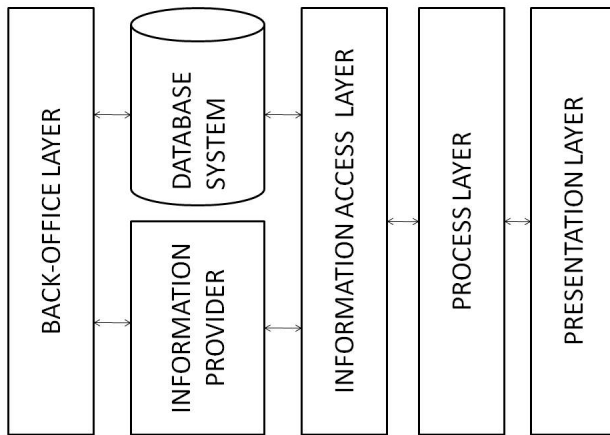


Figure 1: Client's initial self-care system

Basically the client's platform is built in a 3 layered schema where there is a layer responsible for information retrieval, another one to implement business processes and finally a presentation layer responsible for rendering of information on to the web environment.

This design reveals clear concerns from Zon Multimédia architects in allowing easy maintenance and evolution of each of the components.

The first aspect of the designed platform that should be dealt with is the fact that it only supports only one distribution channel (World Wide Web over pc) so in order to support other ways to distribute functionalities it is required to change the paradigm from one distribution channel to multiple distribution channels.

The overall aspect of the platform should be:

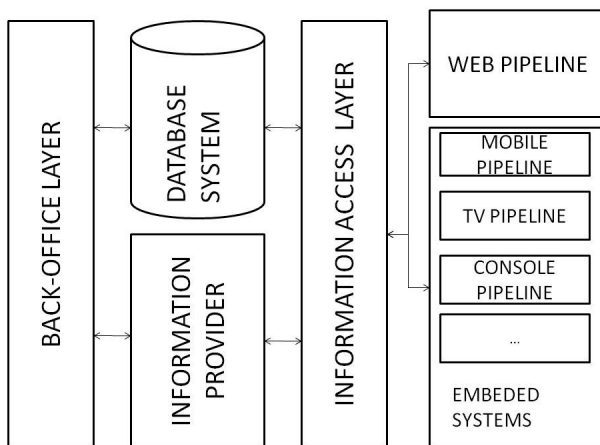


Figure 2 : Architecture Proposal for new self-care environment

This proposal will allow supporting several ways of functionality distribution allowing then to build a truly networked environment where you can manage your client portfolio from the phone, request movies from your web enabled PMP, check your product spending on you TV (these are some of the recommendations).

The target embedded systems for this project and by request of ZON Multimédia are the portable multimedia phones that support http1.0 and the set-top boxes currently in the client's homes.

4. USAGE MODELING

The functionalities elected to be used on the embedded systems was done with some caution because the usability paradigm is quite different from the regular web environment, and may even differ between different types of devices. In order to do the choice, it was identified the need to evaluate what are the most used functionalities by the users. Considering that the ZON Multimédia has a base of 500k net clients, so a sample of 1k clients was used to answer a set of questions, oriented to identify what are the usage patterns of self-care solution.

The presentation layer follows a structure in where exist a first login, followed by a dashboard page (identified as the cockpit), followed by a highly horizontal structured of a list of functionalities.

The cockpit page is a dashboard of predefined overviews of several functionalities (chosen by the user) applied to the client portfolio (Traffic consumption, email management, institutional news, E.P.G. favorites, payment information, etc ...).

The overall structure can be schematized in the following Figure.

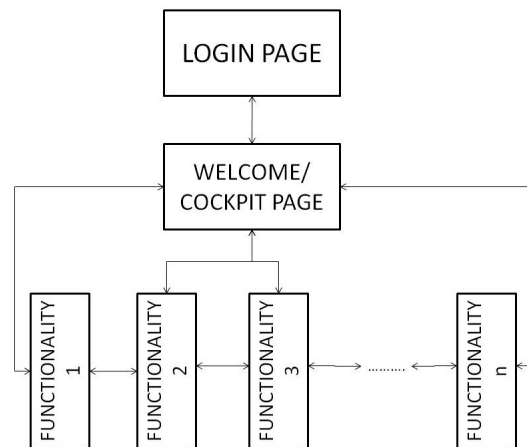


Figure 3: Self-Care simplified navigational diagram

It was revealed by enquire, that the functionalities present on the cockpit represent 80 percent of the most used portal functionalities. When enquired on the possibility of having the same functionalities using a different interface such as an embedded device, the clients revealed high expectations on the simplicity of the interface, fast access times to the information and navigation that requires the least effort possible to access the information.

5. EMBEDDED SYSTEM INTERFACE DESIGN

5.1 Mobile Devices

Designing for Mobile devices required that the architect to consider the following:

- Target platforms identification
- Average screen size admitted
- Content weight on connections and device memory
- Client Application profiling

In the context of the project it was devised that:

- The target is all Html enabled devices (with web browser), this allowed broader range of devices to be supported
- Average size of 200px width by whatever height the page requires, allowing vertical navigation and stretching contents in the horizontal plane.
- Design pages no bigger than 20k in order make them small enough to have faster loading times on the device
- Always have a navigational buttons allowing fast navigation between sections
- Build clear connected navigational graph for simple and clear navigation on the user side

5.2 TV Dependent Equipments

Designing for TV dependent devices required that the architect to consider the following:

- Color usage based on flickering avoidance, contrast enhancement and taking in account older display systems.
- Interface design based normal reading model for TV (Z model reading), and large and simple interface elements.
- Fast loading times
- Easy navigational elements, based on vertical design of pages, least effort needing use the equipment interface (STB remote, videogame console control pad, etc ...)

Based on this, it was devised:

- Avoid flickering colors, pushing a simple and contrasting interface.
- Low weight pages, using a few number of elements to be loaded in order to avoid slow load times
- Build clear connected navigational graph for simple and clear navigation on the user side

6. TEST DESIGN

The test design take in account two stages:

- Pre-launch: where tests look for interface errors and functional errors that are evaluated based on test chart were the functionalities are listed
- Post-launch: these tests aim to evaluate click path and cross information between usage and available

functionalities, allowing detection of architectural flaws in the click path designs

7. CURRENT PROJECT STATUS

Currently the project has the web pipeline and mobile devices pipeline currently working and on production environment.

Platform usage evaluation allowed noticing that:

- Near 67% of users followed all the possible graph navigations, visiting all the available options.
- Average session time is 2,5 minutes were the most used functionalities are E.P.G. listings and payment information
- Client test groups identified 60% of the functionalities as being useful, and easy to use
- From the post mobile device launch 44% of the enquired person question when the TV version would be available for them to manage many things such as premium channels subscription, VOD request management, global payment management and more detailed EPG information.

At this stage of work it also possible to acknowledge that involving the decision makers on step1 and step2 of the process proved to be a profitable decision because it made the decision makers much close to clients desires and even made them more aware of the impact that some devices have on clients life, and by so providing them a better view of embedded devices reality. (For mobile devices)

The above described preliminary results are quite promising exposing the following relevant information:

- Building short click paths in the navigation design, proved to reflect a site navigation that had good acceptance and will also allow to segment marketing strategies when selling advertisement spaces over the pages.
- Early identification of most used functionalities, allowed a correct identification of the application context and by so improving the acceptance in the embedded device users.
- The development process as a viable one to use in a corporate environment.

The TV dependent devices interface is currently being built.

8. ACKNOWLEDGMENTS

My thanks to all the coworker's on Novabase Digital TV and all the decision makers on ZON Multimédia that helped making this project a reality.

9. REFERENCES

- [1] Ding Regan,T. and Todd (2004) Media Center Buddies: Instant Messaging around a Media Center, In Proc. Nordi CHI 2004 141-144

- [2] Nielsen Jakob 2000 Designing Web Usability: the practice of simplicity. New Riders Publishing, Indianapolis ISBN: 1-56205-810-X
- [3] Nielsen Jakob. <http://WWW.useit.com>
- [4] Microsoft Corporation, <http://www.microsoft.com/Mediaroom>
- [5] Opentv Corporation, <http://Opentv.com>
- [6] Springett Mark, Griffiths Richards Accessibility of Interactive Television for Users with Low Vision: Learning from the Web EuroITV 2007 ACM
- [7] Nguyen, Hung Q, Johnson, Bob Hackett, Michael Testing Applications on the Web: Test Planning for Mobile and Internet-Based Systems 2003 Wiley Publishing
- [8] Yates, J. Frank Decision Management – How to assure better decisions in your company Michigan Business school 2003 Josey Bass Printing
- [9] Park, Glen M. Team Players and Teamwork New Strategies for Developing Successful Collaboration 2008 Wiley

NoTube – Making TV a Medium for Personalized Interaction

Lora Aroyo
Vrije Universiteit Amsterdam

Alex Conconi
TXT Polymedia

Stefan Dietze
Open University

Annelies Kaptein
Stoneroos Interactive Television

Lyndon Nixon
STI International

Christoph Nufer
Institut für Rundfunktechnik

Davide Palmisano
Asemantics

Luca Vignaroli
RAI Radiotelevisione Italiana

Milena Yankova
Ontotext Lab, Sirma Group Corp.

contact: l.m.aroy@cs.vu.nl

ABSTRACT

In this paper, we introduce NoTube's vision on deploying semantics in interactive TV context in order to contextualize distributed applications and lift them to a new level of service that provides context-dependent and personalized selection of TV content. Additionally, lifting content consumption from a single-user activity to a community-based experience in a connected multi-device environment is central to the project. Main research questions relate to (1) *data integration and enrichment* - how to achieve unified and simple access to dynamic, growing and distributed multimedia content of diverse formats? (2) *user and context modeling* - what is an appropriate framework for context modeling, incorporating task-, domain and device-specific viewpoints? (3) *context-aware discovery of resources* - how could rather fuzzy matchmaking between potentially infinite contexts and available media resources be achieved? (4) *collaborative architecture for TV content personalization* - how can the combined information about data, context and user be put at disposal of both content providers and end-users in the view of creating extremely personalized services under controlled privacy and security policies? Thus, with the grand challenge in mind - to put the TV viewer back in the driver's seat – we focus on TV content as a medium for personalized interaction between people based on a *service architecture* that caters for a variety of content *metadata*, *delivery channels* and *rendering devices*.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services

General Terms

Design, Human Factors.

Keywords

Interactive Television, personalization, user modeling, context modeling, service oriented architecture, multimedia services.

1.INTRODUCTION

Back in 1960 – there was only one TV at home. It was bound to the living room where everybody would watch the same limited number of linear programs on a small set of channels. This classic TV notion is forever gone in 2010. The TV content has left the living room and has become dissociated from the TV tube. It has moved into the world of Internet and mobile technology. TV content is there in the train, at the back seat of the car, while sitting on a bench in the park. The available TV content, over the IP and broadcast channels, in terms of number and diversity has grown exponentially. The notion of channels has changed drastically: providers just offer a set of live and on-demand programs. 'Zapping until you drop' is replaced by services that enable you to generate your personalized schedule of TV programs. In terms of technology infrastructure there is no difference between TV and computer: the TV content and other streaming content live in a shared connected online world.

The advent of digital broadcasting technology for audio and video content has increased the amount of content from which a radio or TV user can choose dramatically. At the same time, people's preferences have become very much individualized, personal time constraints are complex, and the availability of digital storage technology at the recipient's side has opened a new degree of freedom. From this perspective context-awareness is an essential characteristic of an infrastructure that enables personalized TV service. Modeling and processing of contextual information about consumer preferences and behavior, about capabilities, performance and availability of devices and audiovisual streams is a core research field tackled within the NoTube.

Additionally the NoTube project will take into account community aspects (e.g., how one person can benefit from the media memories of others, or with whom they share common interests) in order to lift content consumption from a single-user activity to a community-based experience. These personalized and shared experiences typically take place in a connected multi-device environment. Also, the traditional of "device" in terms of media channels (e.g., telephone line, internet line, broadcast line) is rapidly disappearing, as the technological infrastructure is moving towards a single-cable concept. The role of these devices

then becomes what they actually should be, namely a vehicle for rendering in an optimal way content for a particular context.

In the landscape of TV content delivery we have to consider three types of parties:

- Broadcasters who provide live and/or on-demand programs.
- Telecommunications companies that bring TV content to consumers.
- TV viewers, who watch TV programs and who interact with the content and with other users of the same content.

The TV content data, the delivery channels and the consumer devices are of a varying nature and cannot be fixed in advance. For example, considerable variations exist in TV content metadata. Availability and popularity of delivery channels is very much dependent on the region. The consumer may use different rendering and interaction devices, such as a stationary TV with remote control, a handheld, or a laptop computer. Transparent development of those personalized services requires middleware that provides a layer of abstraction on top of these varying metadata formats, channels and devices. One of the key challenges of NoTube is therefore the development of a service architecture that can act as such a middleware. Figure 1 illustrates the NoTube vision of the complementary integration of Web, TV and mobile spaces by this middleware in order to enhance the TV experience.

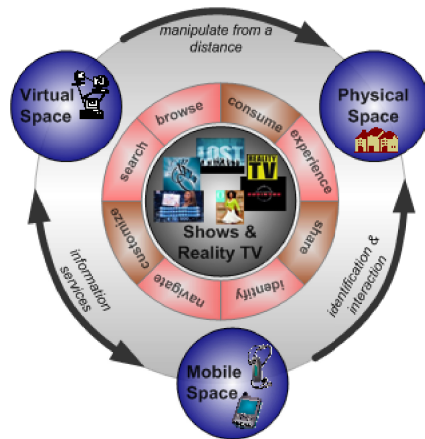


Figure 1: Complementary integration of Web, TV and Mobile spaces to enhance TV experience

In this paper, we introduce the core objectives of the project (Section 2), outline the envisioned service-oriented architecture (Section 3), describe the guiding use cases (Section 4), list the principle challenges (Section 5) and identify the target results (Section 6). In conclusion we discuss the contribution of NoTube to a new innovative system enhancing Interactive Television with advanced personalization technology.

2. CORE OBJECTIVES

We aim at a *user-centric approach* to investigate fundamental aspects of consumers' content-customisation needs, interaction requirements and entertainment wishes, which will shape the

future of the "TV" in all its new forms. "New technology is transforming the TV industry", says Mark Thomson, BBC CEO for Observer. Watching TV more and more happens together with PC-related activities, e.g. chatting with friends, talking on the phone, searching on the Internet for related info about programs. Already in 2000, the trend was by "*couch-and-mouse toys* serving 27 million Americans who were able to watch TV and surf the Web at the same time – the "telewebbers". Nowadays, digital video recording software provides the facility to "time-shift" live television programming through a PC (e.g. via TiVo and ReplayTV). Companies are attempting to bundle "an electronic program guide into its software, along with personal viewing agents that can recommend broadcasted programs based on your viewing habits"¹. For example, while watching "Dancing with the stars", you can vote for your favorite couple, read their biographies, find other programs this week where they appear, record your favorite dances and send them to friends; you can also join live chats during the program. However, "such interactive services stay TV-centric (in the device sense of the term TV) and use Internet and digital technology only to enhance the entertainment that you expect to see when you plop yourself down in front of the tube."²

Hence, our core objectives are developing semantics-based and context aware tools and services for (1) *personalized content selection*, (2) *packaging of content with additional information* (e.g. relevant Web info on program subjects), and (3) *social interaction in/consumer communities*. With semantic based we mean that we build on information integration and representation techniques developed within the Semantic Web area, enabling us to reason about related concepts in different information sources. Context-aware implies that the services are dependant on individual aspects of time, space, task and mood.

3. SERVICE-ORIENTED ARCHITECTURE

Service Oriented Architecture (SOA) is an emerging design model for distributed applications that clearly simplifies the development and re-arrangement of new and existing services. Applied technology for the realization of SOA are Web services, based on SOAP, UDDI and WSDL, that have the big advantage that they are platform independent and based on standard Internet protocols. To cope with the communication and coordination challenges implied by distributed systems of services a Semantic Web Services (SWS) framework is developed as a communication and coordination middleware that allows for decoupling applications in time, space, and reference. Moreover, due to implicit SWS technologies such as mediation, semi-structured data models and knowledge entailment, they also provide means to at least partly decouple applications in vocabulary, i.e. the terms and protocols applied in interactions. One of the primary aims of the NoTube project is to enhance the existing SWS technology to ensure a more adaptive, proactive and context-aware middleware built on top of established SWS frameworks such as WSMO.

To realize the vision of the core objectives we aim at a **service-oriented architecture** with four corner stones: (1) novel

¹ Control your TV with your PC, by Tom Springs, PC World, April 28, 2000

² Analysis: Interactive TV brings new tricks to an old box, by Tom Springs, PC World November 27, 2000

deployment of existing *methods for information integration* in the combined TV-Internet environment; (2) extensions of existing *user and context modeling techniques* to meet the demands of the distributed world; (3) development of *novel reasoning services* for personalized content recommendation; and (4) integration of community-oriented software and social interaction tools, such as chat channels.

Semantic Web Services enable the automatic discovery and selection of distributed resources - services and data exposed via Web services - for a particularly expressed user goal. Current results of SWS research are available in terms of reference ontologies, such as OWL-S and WSMO, as well as comprehensive frameworks (see DIP project results³). Particularly with respect to personalised delivery of media content, dynamic allocation of services, data and metadata is an important requirement to deliver the most appropriate media services to the end user. Context-adaptive delivery of broadcasting media has to consider two distinct context abstractions – i) *the provider context*, where broadcasts have to be directed to classified consumer categories, and ii) *the user and device context*, where media content is adapted to certain device and end-user criteria, such as supported resolution, formats or the preferred interactive feature add-ons. Hence, content adaptation tasks are usually carried out through a sequence of Web service invocations. In that, the *Semantic TV Resource Broker* will enable the dynamic discovery and orchestration of the most appropriate services for a given context.

4. DRIVING USE CASES

The work is steered by three **visionary use cases**, each addressing different dimensions of personalized TV-content interaction, including individual viewers and communities of viewers as well as multi-lingual and multimodal interaction:

- Personalized Semantic News, which focuses on the design and development of a system for the creation of a set of local personalized news services, able to acquire news items from generic broadcast streams, understand the meaning of video news items, understand the physical context in which news items are going to be shown, apply criteria for matching the user profile with the available news items.
- Personalized TV Guide with Adaptive Advertising, which focuses on context identification technologies (e.g. RFID/NFC tags, mobile phones, sensors), as well as different control interfaces (e.g. traditional TV remote control, Web-enabled remote control, mobile device actions) to offer a new Program Guide experience to the consumers; End-to-End Personalized Advertising creation, manipulation, delivery and consumption of personalized advertising messages, also allowing user-generated content and consumer participation to the content value chain.
- Internet TV in the Social Web, which focuses on next generation Web 2.0 technologies and social software to improve interactivity, participation and empowerment of TV consumers to become active protagonists of the TV value chain.

More generally, we will experiment with using media-based events in combination with semantics from the metadata in order

to enable users to find content of a given type, with certain characteristics, which they or related users watched around the time of another media event. The reasoning infrastructure will help to overcome a number of limitations in current Electronic Program Guide (EPG) and move them to Personalized Program and Service Guides. Consider, for example, how to search for programs that are: (1) non-fiction, (2) produced between 1989 and 1995, (3) involve locations in Eastern Europe. Criterion 1 is resolvable using current EPGs. Criterion 2 is not generally supported by current EPGs, but is possible if the program information has been properly populated and indexed along a given timeline. If this is not the case, content analytics and data extraction techniques must be employed to generate the correct information. Criterion 3 again requires appropriate metadata, and also requires a sophisticated knowledge model that is able to represent and reason about part/whole relations between places, countries and regions (e.g. Sofia is in Bulgaria; Bulgaria is in Eastern Europe). The application of shared semantic spaces as medium for information integration of the multi-media system allows for combined data sources and reasoning within arbitrarily established communities.

We intend to improve the quality of News services available to end TV consumers not only for young generations, mobile and Internet surfers, but also for those categories of users who are fond of “traditional” sofa-TV (so-called “lazy” consumers) and not particularly ICT-literate (so-called “TV remote control” clickers). Instead of leaving to the final consumer all the decisions and all the personalization levers (e.g. accessing an EPG and programming by himself his/her own PVR), we propose the possibility to have a provider-consumer collaborative platform, so that part (or most) of the personalization activities are performed at providers’ side on the basis on aggregated privacy-preserving semantic user-context-content profiles, allowing the most lazy users to placidly sit in the sofa and be able to access an highly personalized and interactive News service. Usually the TV consumer accesses to the entire broadcasted news program recorded, by using EPG information; this fact raises problems about news program EPG identification, about a huge memory space needed to record all the content and about the access time to a single news item of interest which may be very long.

5. RESEARCH CHALLENGES

The ICT landscape is developing into a highly-interactive distributed environment in which people interact with multiple devices (e.g. portable devices such as mobile phones and home equipment such as TV’s) and multiple applications (e.g. computer programs such as Web browsers and dedicated Web services). The most recent proof for this is ‘technology embraced by public’ as Darren Waters calls it. The author shares the observations of Sean Wargo, director of industry analysis for the Consumer Electronics Association (CEA), at the Consumer Electronics Show 2007 in Las Vegas, one of the world’s largest technology shows. People more than ever become early adopters of technology. Globally the industry is being driven by the shift away from old models - from physical to digital. New methods emerge for getting content such as TV programs via the Web. Almost half of the people want to watch TV content on their PC’s. They want to make a bridge between a TV and a PC, perhaps even sitting in a home office. As we can see, technologies in these fields are rapidly progressing, but the user is lost. The

³ DIP Project: <http://dip.semanticweb.org>

information overload is enormous and the content presented is hardly adapted to the prior knowledge, to the preferences and to the current situation of the user. Not only the users, but also the industry comes to the realization that content services and user experience are becoming crucially important.

Additionally, users are increasingly involved in multiple virtual environments (e.g. MySpace, Flickr, YouTube, Amazon, entertainment sites) in each of them with a different identity (e.g. login information, preferences). There is very limited integration of these user profiles, or if there exists integration it is not always under the control of the user and there is a lack of transparency in the use of personal data between different applications. As most of those services are relatively new and still aim at gaining critical mass of users, there is still not a methodological approach of how to assess the users experience and improve in an evolutionary way the provision of the services. Personalization in information retrieval and information presentation [3] has therefore become a key issue. Successful personalization experiments have been done, in e-commerce [1] and news websites [2], where the most common example is the Amazon.com recommendations. Personalization is seen as a key ingredient of the so-called “Web 2.0” applications [10,14]. However, such personalization is still local. In the example: the personalized information is only valid for Amazon; it cannot be used for other information services, nor can Amazon cater for different “modes” of a user, e.g. when looking for something personal, or when looking for a gift for his/her kids or friends.

This new ICT landscape can be characterized with three key terms: Distributed Interactive, and Multi-device. In the project we use the acronym DIM to refer to this application setting. In DIM environments, individuals and groups make use of multiple devices in multiple distributed contexts (e.g. entertainment, personal information management). This results in complex interaction patterns requiring integrated views of distributed data collections, multiple modelling perspectives of content, user and environment data, as well as an increased need for personalized information presentation. To realize distributed personalization in DIM environments, we identify four main challenges: (1) *data integration and enrichment* - how to achieve unified and simple access to dynamic, growing and distributed multimedia content of diverse formats? (2) *user and context modeling* - what is an appropriate framework for context modeling, incorporating task-, domain and device-specific viewpoints? (3) *context-aware discovery of resources* - how could rather fuzzy matchmaking between potentially infinite contexts and available media resources be achieved? (4) *collaborative architecture for TV content personalization* - how can the information about data, context and user be put at disposal of both content providers and end-users in the view of creating extremely personalized services under controlled privacy and security policies?

To realize personalized TV in DIM environments we need to address the challenges outlined above. For the data integration problem there are already a number of general working solutions within the area of Semantic Web [11,12,13,6,7]. We plan to select appropriate solutions and deploy them in our project setting. Thus, data integration will not be the main focus of research in this project. User modelling is an area where multiple research examples, primarily in e-learning, have been presented [4,5]. However, the field still lacks concrete standardization of the user and context representation for both local and distributed

applications [8]. Similar is the situation with personalized presentation of content, where the research is still in early stages [9]. The work here targets the use of open standards and reaching consensus for user and context modelling and personalized content presentation in DIM environments.

6. DIM ENVIRONMENTS

In the DIM environment we can perceive three “spaces”, namely the physical, virtual and mobile spaces, see Figure 1. The user resides in the *physical space*, e.g., at home, on the street, in a museum. This space encompasses task-specific devices such as TV, VCR and information displays in museums. The user’s prime goal in interacting with those devices is to consume and experience content. The rich-interaction environment of the Web plays the role of the *virtual space*, where users typically work with multiple domain-specific applications and perform information intensive tasks, such as searching, browsing and navigating in various content collections. The information prepared on the Web could be further used in the physical space: preferences for TV programs could be uploaded to a TV set-top box. The *mobile space* encompasses all mobile devices (iPod’s, PDA’s), each containing a fragmented portion of user preferences and data. In each of the spaces one can interact with content: one can physically look at a movie on the TV, view a trailer of it on a website or look at a picture sms’ed to your mobile phone. In earlier work we have identified six types of user interaction with content: search, browse, navigate, consume, experience, and share. We have also identified two additional user interactions, which operate on user data: namely identify and customize. All these interactions can be performed in each space. However, not all spaces provide the most efficient means for the user to realize the tasks easily. To enhance the user’s experience and consumption of content in the physical world, we integrate the three spaces in such a way that the user can perform most efficiently the interactions in the dedicated spaces and use the results in the complementary ones. For example, identifying a user can more easily be done through a mobile phone. Browsing and searching is easier with a computer keyboard and mouse than with a TV remote control or a phone keypad.

7. CONCLUSION

In this paper we have presented the NoTube project and its vision for making TV a platform for personalised interaction. It will be possible to follow the activities and achievements of NoTube at the project website <http://www.notube.tv>

REFERENCES

- [1] Ardissono, L., Console, L., and I. Torre. An adaptive system for the personalized access to news, 2001.
- [2] Ardissono, L. and A. Goy. Tailoring the interaction with users in web stores. *UMUAI*, 10(4):251–303, 2000.
- [3] Brafman, R.I., C. Domshlak, and S. E. Shimony. Qualitative decision making in adaptive presentation of structured information. *ACM Trans. Inf. Syst.*, 22(4):503–539, 2004.
- [4] Chin, D.N. Empirical evaluation of user models and user-adapted systems. *UMUAI Journal*, 11(1):181–194, 2003.

- [5] Fink, J. and A. Kobsa. A review and analysis of commercial user modeling servers for personalization on the WWW, 2001.
- [6] van Gendt, M., A. Isaac, L. van der Meij, and S. Schlobach. Semantic web techniques for multiple views on heterogeneous collections: a case study. In ECDL 2006, pp 426–437, 2006.
- [7] Klein, M. Combining and relating ontologies: an analysis of problems and solutions. Workshop on Ontologies and Information Sharing, 2001.
- [8] Kobsa, A. Generic user modeling systems. User Modeling and User-Adapted Interaction, 11(1-2):49–63, 2001.
- [9] Light, M. and M. T. Maybury. Personalized multimedia information access. Commun. ACM, 45(5):54–59, 2002.
- [10] Millard, D. and M. Ross. Web 2.0: hypertext by any other name? In HYPERTEXT '06, 27–30, 2006. ACM Press.
- [11] Noy, N. and M. A. Musen. Prompt: Algorithm and tool for automated ontology merging and alignment. Workshop on Ontologies and Information Sharing, 63–70, 2001.
- [12] Schreiber, G., et al, MultimediaN E-Culture Demonstrator. In The Semantic Web – ISWC 2006, Athens, Georgia, volume 4273 of LNCS, pages 951–958. Springer Verlag, November 2006. Winner Semantic Web Challenge 2006.
- [13] Shvaiko, P., and J. Euzenat. A survey of schema-based matching approaches. J. on Data Semantics, 146–171, 2005.
- [14] Webster, D., et al. Context-orientated news filtering for web 2.0 and beyond. In WWW '06, 1001–1002, 2006. ACM Press.

Modela-TV: Service Personalization and Business Model Management for Mobile TV

Francisco Fraile

Multimedia Communications Group
iTEAM, Polytechnic University of
Valencia (UPV), Spain
Phone: 34-963879717

ffraile@iteam.upv.es

Ismael de Fez

Multimedia Communications Group
iTEAM, Polytechnic University of
Valencia (UPV), Spain
Phone: 34-963879717

isdefez@iteam.upv.es

Juan Carlos Guerri

Multimedia Communications Group
iTEAM, Polytechnic University of
Valencia (UPV), Spain
Phone: 34-963879717

jcguerri@com.upv.es

ABSTRACT

This paper presents Modela-TV, a research project which investigates how free-to-air file download services can be the basis of new business models in mobile multimedia content delivery networks. Hereby, we present the proposed system architecture, together with implementation guidelines for the main technologies involved. A novel method for the personalization of broadcast file download services, based on cache management, is introduced. This represents one of the main outcomes of the project, as personalization is considered a key feature in mobile TV services.

Categories and Subject Descriptors

D.2.6 [Software Engineering]: Programming Environments - *Interactive environments*. E.5 [Files]: *Optimization, organization/structure*. H.3.3 [Information Search and Retrieval]: *Information filtering*.

General Terms

Design, Performance, Management.

Keywords

Mobile TV, Service Prototype, Personalization, Filecasting, Content management, Service Guide.

1. INTRODUCTION

Current Mobile TV revenue streams are primarily based on billing customers, either for content and transmission, as in the case of On Demand Streaming Services deployed by 3G Mobile Network Operators, or for access to a bouquet of premium TV channels, as in the case of Mobile Broadcast Services. In the traditional (non-mobile) television business, main revenue streams are generated by advertisement on free-to-air services, but also by premium and pay-per-view based business models. Enabling such a scenario in the Mobile TV landscape is regarded as a key factor to boost the penetration of mobile TV services. Current revenue streams are too weak to meet the expectations of all the agents in the value chain. A prove of this fact is that most of the current incomes of mobile streaming services are held by content providers.

On the other hand, the availability of content for download on the Internet has shifted user habits and expectations on media services. Regardless of legal implications, users regard content as easily available and affordable (if not free). In a mobile scenario,

this implies that users expect receiving content of their personal interest during the short time they are connected to the service.

For these reasons, the project "Service Personalization and Business Model Management for Mobile TV (Modela-TV)" proposes a Mobile TV service which allows service operators to use different business models (free-to-air, pay, premium...) simultaneously when offering content to mobile users. The service takes into account user expectations with regards to content availability and cost. The aim of the project is to develop a business model management platform for mobile multimedia delivery services through which, a) service operators can enable several revenue streams simultaneously and b) users can access personalized content at an affordable cost.

2. TECHNOLOGIES

The Modela-TV platform manages several technologies but relies mainly on two: on one hand, FLUTE, for the reliable distribution of files over the mobile broadcast/multicast network. On the other hand, the platform also manages a DRM subsystem in order to protect the intellectual property of the content to be delivered. These systems provide an adequate technical framework to deal with the commercialization of protected content. New ways of delivering content are introduced, which are meant to adapt the usage of the broadcast network in order to match the expectations of the users.

2.1 FLUTE

FLUTE (*File Delivery over Unidirectional Transport*) [1] is a protocol for the unidirectional delivery of files over the Internet, which is particularly suited to multicast networks. The protocol uses ALC (*Asynchronous Layered Coding*) for a reliable and scalable transport of the multicast files. Furthermore, ALC implements a Forward Error Correction block to accomplish reliable multicast.

A FLUTE session consists of a set of logically grouped ALC channels associated with a single source sending FLUTE packets. Each FLUTE session is uniquely identified by the source IP address and the Transport Session Identifier (TSI). A receiver joins the channel to start receiving the data packets sent by the source. Previously, receivers must know the multicast addresses of the channels using out-of-band mechanisms such as SDP (*Session Description Protocol*).

There are two kinds of FLUTE sessions, File Delivery Sessions and File Carousels. The latter consists of a cyclical transmission of all files to be transmitted on a seamlessly endless loop and it is the one normally used, since cyclical transmission provides a

reliable mechanism to receive files correctly. Furthermore, sessions can be static or dynamic, depending on whether the contents of the session change during its lifetime.

One of the fundamental components of FLUTE is the File Delivery Table (FDT), which provides a means to describe various attributes associated with the files that are included within the session, for example the file *Content-Location* (URL or URI of the item) or its numeric identifier TOI (Transport Object Identifier), used by the ALC client to filter the packets corresponding to the wanted files. The FDT is delivered as FDT Instances, which are XML fragments providing descriptions of one or several files over ALC with TOI = 0. Receivers need parsing the FDT Instance that describes a file before it can start the download.

DVB IPDC (IP Datacast) [2] defines file grouping in FLUTE as a mechanism to download together files that have some kind of dependency at application level. This grouping is performed by means of the elements *Group* contained in the *file* element of the FDT Instance that describes a file. For each *file* element, one or several elements of type *Group* may be defined. The FLUTE clients that start downloading a file with a given *Group* element must start download all files with the same *Group* tag. Thus, file grouping enables an efficient caching of files.

2.2 DRM

The term DRM (*Digital Rights Management*) refers to a series of technologies that enable the consumption of licensed digital content in a controlled environment. The OMA BCAST Service and Content Protection [3] DRM Profile provides such a technological framework in order to support the delivery of licensed content over broadcast networks under a variety of business models. According to these specifications, the usage of DRM in Modela-TV is explained below.

Modela-TV adopts the *Subscriber Group Addressing* feature, which allows to define the exact group of broadcast receivers that will be able to access the protected content. Access to content is granted according to the terms and conditions defined by the content owner for the specific subscriber group (e.g. content preview, advertisement placement). *Subscriber Group Addressing* consists of encrypting ROs (*Right Objects*) with UGKs (*Unique Group Keys*). ROs are messages that express these access terms and provide the means to access the content. Thus, in the registration phase, all authorized receivers in a *Subscriber Group* are provided with the same UGK which will allow them to decrypt a series of ROs.

Furthermore, in Modela-TV the content is encrypted with CEKs (*Content Encryption Keys*) and encapsulated in *dcf* (*DRM Content Format*) files. Upon reception of a *dcf* file, a receiver will use its keys (i.e. a set of UGK) to decrypt the ROs within the *dcf* file. If a RO is decrypted, the receiver will use the CEK to decrypt the content and display it to the user, according to the rules expressed in the RO.

2.3 ESG

The Electronic Service Guide (ESG) is a file download service used to provide users and terminals with a description of the services available on an IP platform (similar to Electronic Program Guides of Digital TV Services).

Modela-TV adopts the usage of OMA BCAST ESG standard [4] as follows: the *Access* fragments of the ESG are used to provide the FLUTE session information for File Download services (referencing a FLUTE SDP fragment). The *Schedule* fragments of a content file instantiate the element *ContentLocation*, which value is equal to that of the attribute *Content-Location* of the FDT Instance of the file, thus completing the access information required by the FLUTE client to fetch the file from the carousel.

Moreover, the *Access* fragments contain a *KeyManagementSystem* element which indicates that the DRM profile is used. The *RightsIssuerURI* attribute provides receivers with information to obtain rights associated to the content item. Additionally, *PurchaseChannel* fragments establish a relationship between the URIs and URLs of RIs, so that receivers can contact RIs if the user wants to acquire new rights.

Lastly, each file of the carousel is associated with a *Content* fragment which provides descriptive information such as parental rating, textual descriptions or TV-Anytime [5] genre descriptions. This descriptive information is used to implement personalization features in the service, as explained later.

3. ARCHITECTURE

The figure 1 shows the general architecture of the system, showing the most relevant technologies involved in the project.

The *Business Model Manager* is the main entity of the architecture and the interface between the platform and the operator. It manages the content ingestion and the Digital Rights associated to each media item. Through the *Business Model Manager*, the operator establishes the relationships of the content items with each Subscriber Group. This way, the management of the subscriber data base, the repository of content and the associated metadata (user profiling, purchase information and content description) are also in control of the operator through the user interface of the *Business Model Manager*. According to the setup provided by the operator, the *DRM subsystem* encrypts the content items and produces the *DRM Content Format* (*dcf*) files which are then aggregated to the free-to-air broadcast by the *Filecast Server*. The latter also delivers the metadata (content descriptions, purchase information, etc.) to users in the service area by aggregating the ESG of the service.

The *Filecast Server* is the entity that deals with the delivery of the content through the mobile broadcast network. The *Filecast Server* queries the *Business Model Manager* to obtain the metadata associated to each content item and aggregates this information in the generated ESG. The ESG is broadcasted and then received by the client application to discover which content items are available and their characteristics. Furthermore, the *Filecast Server* also manages the distribution of the content associated to each service by managing the configuration of the FLUTE carousel sessions and the definition of the associated File Delivery Tables. The software supports the definition of File Groups which, as described in the FLUTE section, can be used to accelerate the acquisition of a compilation of content items.

In the current version, the *Filecast Server* provides a constant bit rate IP flow transporting the FLUTE sessions and the File Carousel Manager is responsible for scheduling the transmission of files. An optimal transmission schedule minimizes the overall access time to files in the service area [6]. In order to accomplish

this, instances of File Carousel Manager can configure the schedule of the transmission on a per FLUTE packet basis, thus enabling the implementation of advanced scheduling policies.

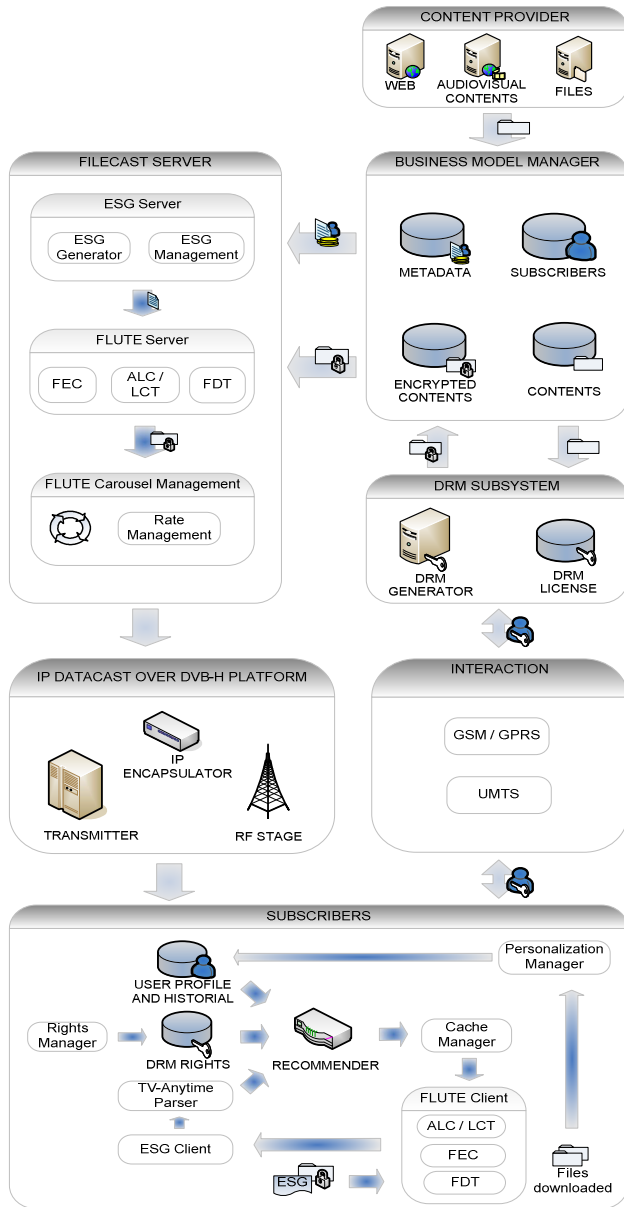


Figure 1. Modela-TV general architecture

In the case of study, the content is delivered by means of a *DVB-H distribution platform*. Moreover, the *Filecast Server* aggregates the FLUTE sessions which correspond to the content delivery services and their Service Guide as a set of IP multicast streams that are encapsulated into an MPEG Transport Stream. The latter is broadcasted by the DVB-H air interface. The IP Encapsulator sets the configuration of the MPE encapsulation (time slicing and MPE-FEC parameters) and aggregates the metadata tables (PSI/SI) which are necessary to extract the IP data from the MPEG-TS by the receivers.

The client application provides *subscribers* with a user interface which discovers and lists all the content available through the service. Once the DVB-H receiver tunes to the frequency that transports the services, the client application receives the Service Guide and presents a list of the available content for download. When the user selects a content item, this is downloaded to the terminal as a background service. The client application also allows entering user preferences and downloading content accordingly. In order to do this, the client application interprets the content descriptions provided through the ESG and selects the media items which better suite the user preferences.

If *interaction* is required, an interactive channel like GPRS or UMTS can be used. As explained, the ESG provides a Right Issuer URL to enable the communication with the system for the acquisition of digital rights, in order to gain access to the desired content. The request and later acquisition of these rights is carried out by this block, since DVB-H does not provide a return channel.

4. PERSONALIZATION

Personalization is one of the most important design aspects of mobile multimedia services [7] and therefore, it represents one of the main goals of the project. Personalization provides many advantages to the user, for instance the automatic discovery of interesting content. The perception of the service adapting to the user preferences and needs encourages a positive experience, most significantly when accessed from personal devices such as mobile phones.

Personalization is also a great asset for content distributors. An efficient and functional application favors the consumption of content, thus increasing the market share of mobile content production. Since users receive content according to their preferences, content distributors can rely on the platform delivering the content only to their target audience and advertisers can improve the impact of their campaigns.

In Modela-TV, personalization is achieved by means of a recommender on top of the DVB-H middleware that interacts with the ESG client and the cache manager as explained hereafter.

First, terminals only store in cache those files for which the user has associated access rights, thanks to the coupling between the DRM subsystem and the *Filecast Server*. This is accomplished by relating the Subscribers Groups to the values of the *Group* elements that are included in the FDT. As explained, the UGKs are used to encrypt Right Objects that target a group of users, whereas all files with common *Group* tags are fetched by the FLUTE client after a single download instruction. So, if a user decides to download a file, the application will first verify that it has access to it. If negative, it will prompt the user to acquire access rights (using the *Purchase Channel* information included in the Service Guide). If positive, the FLUTE client will start downloading that file and all the files with the same *Group* tag, which will be stored in cache.

Accordingly, the average access time to any other file for which the user has associated rights is reduced, since all content items associated to the user's categories of subscriber are cached automatically by the FLUTE client (recall that all files in a Group are downloaded by a single download operation on any of them). If the size of the set of files to cache is bigger than the size of the

allocated memory, then the caching replacement policy will determine which files remain in cache. In Modela-TV the caching replacement policies are based on a method named *PIX* (*Probability Inverse Transmission Rate*) and described below.

In a broadcast scenario, the cache management algorithm [6] must store in cache the objects with higher ratio between the future probability of access, P_i , and the object transmission rate, since the value of storing an object in cache is inversely proportional to its rate. This is due to the fact that, if an object has a long transmission cycle, then the access time of that file is drastically reduced in the event of a cache hit. Thus, for each incoming file, the cache replacement policy computes its *PIX* ratio, compares it to the files stored in cache and discards the file with lower *PIX*.

In order to implement *PIX* based cache replacement policies, an estimation of the local probability of access is needed. In our case, this is provided by the recommendation tool. Recommenders are software tools that estimate the usefulness of content items in a catalogue for a user [8]. In this proposal, it is assumed that the usefulness of a content item is directly proportional to its future probability of access P_i . Upon this assumption, recommendation tools can be used to derive an estimation of the future probability of access to items. This way, for each content item in the carousel, the recommender compares the TV-Anytime descriptions of the content with a user profile that represents the preferences of the user, in order to derive an estimation of P_i . With this information, the cache manager stores in cache memory the files with a higher *PIX* ratio.

The *Personalization Manager* tracks the content consumed by the user, creates user profiles and updates history data so that the recommender can estimate P_i accurately from realistic information about the consumer preferences.

5. CONCLUSIONS AND FUTURE WORK

Modela-TV provides with a platform for the deployment of personalized multimedia content delivery services in mobile multicast networks. The design of the system architecture of Modela-TV has taken into account the expectations of users with regards to personalization, download speed and content availability, without compromising the interests of content license holders. Furthermore, service operators are provided with means to monetize the content through a variety of revenue streams at a very affordable cost, compared to current mobile multimedia content delivery services.

Modela-TV also proposes a new method for the personalization of multicast file download services, based on the use of FLUTE groups and content descriptions to optimize the usage of the cache memory allocated for the service.

A prototype of the platform has been implemented and it has been tested in a test laboratory assembled at the Institute of Telecommunications and Multimedia Applications (iTEAM) of the Polytechnic University of Valencia, as illustrates the figure 2.

In this figure we can see the equipment employed to perform the architecture tests and the running application. The current hardware for the client prototype consists of PDA with a SDIO DVB-H receiver (SIDSA). Application GUI is shown on the screen. Once the client is successfully connected to an IP multicast address, port and TSI, the user can start downloading files and groups in the carousel.

Beside the laboratory tests, the service is going to be tested by means of a pilot transmitter installed in the University Campus. The Radio Frequency Planning of the test transmitter has been performed with iTEAM's planning tool, guaranteeing coverage in all the University Campus. The pilot will host field trials and end user trials.



Figure 2. Modela-TV Application

Current research is focused on the study of optimal file schedule configurations that minimize the access time to files. Also, the project will evaluate how the design of the recommender affects the performance of the cache manager.

6. ACKNOWLEDGMENTS

This project has been financed since 2007 by the Spanish Ministry of Industry, Tourism and Commerce, within the program "Avanza I+D", as part of the financing of the National Plan of Scientific Research, Development and Technologic Innovation. The project has also been supported by Interactive TV Arena, in the framework of the co-operation agreement held with the iTEAM.

7. REFERENCES

- [1] Paila, T., Luby, M., Lehtonen, R., Roca V., Walsh, R.: File Delivery over Unidirectional Transport. RFC 3926 (2004).
- [2] ETSI, TS 102 472 v1.2.1: Digital Video Broadcasting (DVB); IP Datacast over DVB-H, Content Delivery Protocols" (2006)
- [3] Open Mobile Alliance: Service and Content Protection for Mobile Broadcast Services V.1.1 (2008).
- [4] Open Mobile Alliance: Service Guide for Mobile Broadcast Services, Candidate Version 1.0 (2008).
- [5] ETSI, TS 102 822-3-1 v1.4.1: Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems ("TV-Anytime") (2007).
- [6] S. Acharya, R. Alonso, M. Franklin and S. Zdonik: Broadcast disks: data management for asymmetric communication environments, ACM SIGMOD Conf., 1995, pp. 199-210.
- [7] K. Chorianopoulos. Personalized and mobile digital TV applications. Multimedia Tools and Applications, Springer, 2007.
- [8] G. Adomavicius, A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-Of-The-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6 (June 2005), pp. 734-759.

An Optimized Framework for Interactive TV Services Broadcasting

José Antonio Hurtado-Dueñas
jahurtado@citic.es

Bernardo Ruiz-Villalobos
bruiz@citic.es

Rocío Padilla-Conejo
rpadilla@citic.es

J. R. Salinas
jsalinas@citic.es

Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones (CITIC).
6 Marie Curie, Parque Tecnológico de Andalucía, 29590 Campanillas (Málaga), Spain. +34 952028610

ABSTRACT

Great research efforts have been focused on Digital Terrestrial Television (DTT) technology in the last years. The digital modulation and codification features of DTT have made it possible to offer new interactive services and applications. The ability of researchers and producers to develop and optimize these new applications in a fast and reliable manner to cover the new users' needs will determine the future of DTT. Several standards and middleware platforms have been developed to build interactive applications. An example of such a platform is the Multimedia Home Platform (MHP), which is established in most European countries. In this paper, a new MHP based optimized framework to develop and broadcast interactive television applications is presented. The system has been successfully tested and compared in terms of time and size efficiency against another platform belonging to a Spanish state-owned television channel.

Categories and Subject Descriptors

D.3.3 [Programming Languages]: Language Constructs and Features – *Frameworks, modules, data types and structures.*

General Terms

Performance, Design, Reliability, Human Factors.

Keywords

Digital Terrestrial Television, interactive applications framework, Multimedia Home Platform.

1. INTRODUCTION

Digital Terrestrial Television (DTT) is the new technology that will replace the traditional analog television. It is based on the use of digital encoding and modulation to manage audio and video contents, achieving a more efficient bandwidth use.

Besides the audio and video content, another kind of digital information can be sent to provide new services and applications that might run on the users' receivers or set-top-boxes. The Multimedia Home Platform (MHP) standard [4] [9] has been chosen to develop these applications in most countries in Europe.

The new DTT services and applications provide very interesting interactive features. The user does not have a passive role anymore as he can freely change the contents and information he is receiving from the television. Basically, there are two different types of interactivity: (a) local interactivity, which is related to the

users that may interact locally with data stored in the set-top-box, and (b) real interactivity, that allows the user to provide response data and obtain data by using a return channel.

DTT Technology also helps to reduce the problem of the digital gap, which refers to the difference in opportunities for development between those who have access to new communication technologies and those who do not, thanks to its universality, set-top-boxes low cost and ease of learning.

Due to the DTT and MHP market growth, an efficient and robust application development framework becomes a key production factor to become competitive and allow fast implementation of new interactive services and functionalities.

2. STATE OF THE ART

As an initiative of the Digital Video Broadcasting Project (DVB), the Multimedia Home Platform (MHP) defines a generic interface between interactive digital applications and the terminals on which those applications execute. This interface de-couples different providers' applications from the specific hardware and software details of different MHP terminal implementations. The terminal can be a set-top-box, a television set or a desktop computer. Therefore, MHP envisions a truly horizontal market in contents, applications and services over multiple delivery systems including cable, satellite, terrestrial broadcasting and even third generation mobile phones.

Other standards for interactive TV around the world are the OpenCable Applications Platform (OCAP) [13] that is the software environment standard developed by the North American cable TV industry to enable interactive applications to run on cable and the Multimedia-Hypermedia Experts Group (MHEG) [10] that consists in an open standard interactive TV designed to provide interactive services in UK. Different transmission systems may adopt MHP middleware through the Globally Executable MHP (GEM) standard [3], which is a core of MHP APIs, where the DVB-transmission specific elements were removed.

DVB defines a suite of APIs that includes most of the Java TV™ API [17], HAVi (user interface) [5], DAVIC APIs [2] and DVB APIs. The applications downloaded to the set-top-box are Java™ applications called Xlets, built on a suite of APIs, tailored specifically for the interactive TV environment.

In [18], a browser approach for efficient and rapid deployment of MHP applications based on HTML/XML resources is presented. The first MHP browser based on DVB-HTML was Pontegra, developed by Nionex [12] in 2002. Commercial tmplayer from

Tmira [19] and open source Yambo developed by Cineca [1] are also examples of recent MHP browsers that rely on XML descriptions to build simple applications. In this paper, we present a new optimized framework to develop and broadcast interactive TV applications based on a similar approach and focused on achieving a proper performance in terms of time and size.

3. PROPOSED FRAMEWORK

3.1 Architecture Overview

The proposed framework is shown in Figure 1. This framework introduces an additional software layer on top of the MHP layer in order to run applications described in XML documents.

This additional layer, known as a “XML Browser”, is a standard Xlet, which can be present with others Xlets without any problem. This browser makes it possible to run a complete interactive application, written entirely in XML format with the possibility of adding new feature extensions through plugins written in Java.

The XML Browser provides a higher level of abstraction for application development making it easy to write several tasks, which are more tedious to write in Java. The largest part of the code needed to perform typical operations with the MHP API is placed in the browser code, which is shared with all applications, thereby saving space in the object carousel [6] and reducing the startup time. The different components in the browser architecture are described in the following sections.

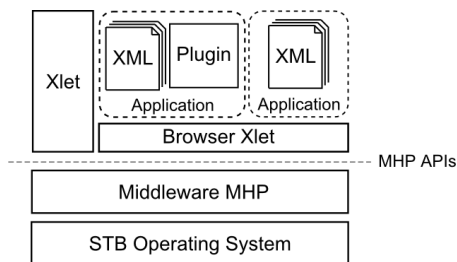


Figure 1. Software Architecture Diagram

3.2 XML Application Description

XML format was chosen to describe applications because it is extensible and there exist xml parsing libraries for Java, such as nanoxml [11], which has a small size and a good performance. In this framework, all applications are described in one or more XML documents which are processed at runtime by the browser.

Using XML for describing applications minimizes the amount of required space in the object carousel. The browser size is around 300Kb, which is rather high for a typical interactive application. However, XML applications are quite small because most of the needed functionality in those applications is implemented as high level functions in the browser code, which is shared by all XML applications. Sharing the browser code between applications also improves the loading time, because most of the needed Java classes are already loaded in the memory of the set-top-box.

XML files can be loaded from the object carousel or from the return channel via http/https protocols. They even can be generated dynamically by an application server. This feature brings the possibility of running server side applications with dynamic contents making it possible to use the TV as a true interface to the World Wide Web.

3.3 User Interface Component Set

The system has a set of graphical components that implement the user graphic interface. The components consist basically of enhanced text using styles, text input for the entry of text on the part of the user, containers, images, menus, buttons, animations, tables, comboboxes, checkbox, and textlists. These widgets are not included in the typical widgets of HAVi.

The XML description of the user graphic interface includes all the information associated with the graphical components set of the proposed framework, defining properties of each component, user events, and actions associated with those. An example of XML description is showed in List 1.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<application>
  <screen id="pulsar_rojo" background="" focus="pulsar_rojo">
    <video x="0" y="0" width="720" height="576" />
    <animation id="botonrojo" x="46" y="500" width="240"
      height="49" time="500">
      <image file="citic/images/pulsar_rojo.png" />
    </animation>
    <keyevent key="VK_COLORED_KEY_0"
      command="exit()" />
  </screen>
</application>
```

List 1. Example of Browser XML code

3.4 User Interface Builder

The user interface builder is the module that takes the XML description files as the input and processes all XML elements in order to build the screens on TV. As an initial step, the XML file is parsed and loaded into the memory of the device. Then, the browser takes the initial screen element and it starts the building process of that screen in order to show it on TV.

The time needed to build a screen depends on the number of components defined in it. If an application has a large amount of screens, processing all of them at the beginning could take a lot of time. For this reason, every screen is built at the moment of being showed, hence speeding up the loading time of the application.

3.5 Command Interpreter

In the proposed architecture, a user event may associate an action to a button on the remote control. The actions are defined by commands with a particular syntax that are processed by an interpreter integrated inside the framework. The commands can be included inside the XML file descriptor to perform a specific action of the component of the application, such as a script.

The command interpreter is also capable of handling variables, used as arguments of the commands and useful for its capacity to work on different screens, menus or components. It may also process scripts, estimate each command and execute all the actions associated to them.

3.6 Event Manager

Every component of the screen and even the screen itself may contain user event handling elements in order to perform a custom action when the user presses a key of the remote control.

The event manager is in charge of processing the user event handlers defined in the XML application description. The event handling process depends on the selected component at the moment of pressing the key. This event handling mechanism covers the requirements for most applications in DTT. In other cases, the event handling can be processed with a browser plugin written entirely in Java code using AWT event listeners.

3.7 Plugin Extensions

The XML browser solves the aspects associated with the MHP graphical interface, although it is not able to accomplish more complex tasks, such as the execution of algorithms or the use of a smartcard. Thus, it is necessary to implement some mechanisms to create hybrid applications that use the browser application to define the graphical interface and its behavior and, on the other hand, other specific elements written on Java from scratch. The proposed system uses plugins to solve the specific components implementation and allows invoking their own methods.

3.8 Security

An internet access profile was introduced in MHP 1.1 that enables the access of the applications to the internet from the set-top-box. This fact brings new opportunities to provide web services through TV. Then it is possible to load XML applications over the return channel via HTTP protocol from an application server.

Some services may require secure access to the application server. In that case, it is possible to make a secure connection using SSL sockets because MHP 1.1 supports them.

DDT receivers often have a smartcard slot, which can be accessed using standard SATSA APDU API [7] in order to store user information as passwords, user settings, encryption keys, etc. XML Applications can be broadcasted in an encrypted form in order to grant access only for users with the smartcard which contains the proper keys for decryption.

4. SYSTEM EVALUATION

In this section, the proposed system performance is compared with another interactive applications platform developed for a Spanish state-owned television channel. In the following sections, we will refer to this platform as “reference system”.

4.1 Software and hardware resources

Several software and hardware resources were used to perform this evaluation. StreamGURU MPEG Analyzer [15] was used to extract files contained in carousels. TSReader [20] was used to obtain elemental streams bitrates from transport streams. iMux multiplexer from MIT-xperts [8] was used for frame generation, while interactive applications were developed using MHP 1.1.2 middleware from Osmosys [14]. Finally, Strong 5510MHP devices were used as set-top-boxes [16].

4.2 Evaluation initial issues

Certain issues related with the following concepts were taken into account before both systems were compared. Global bitrate affects directly to the applications loading time. Thus, it was established to 790.94 Kbps for both set of applications, in order to evaluate both systems on equal terms.

The reference system contains applications that are designed on a traditional Xlet manner, which implies that every application is written from scratch, whereas the applications running in the proposed system are services that run over the browser.

The interval of time in which an application is completely emitted on the broadcast channel is called cycle time. It depends on bitrate and applications size. The proposed system uses a single carousel where the bitrate is distributed in a fair way amongst every single application. Therefore, the same cycle time is achieved for all the applications in the carousel. On the other hand, the reference system uses several carousels with different and specific bitrates, so each application has a specific cycle time.

The applications running in the proposed system were tested, using the three caching approaches described in MHP standard, but no significant differences were found in the tests

Loading time is defined as the interval of time from the moment an application is launched until it can be executed. Increasing the number of applications on the transport stream implies a bigger amount of data and a loading time increase, as bitrate is fixed. On the proposed system, 11 applications are delivered. The browser application is excluded from this number, as it cannot provide any service on its own. The reference system delivers 6 applications, including a launcher application.

4.3 Applications analysis

The applications to be evaluated were previously analyzed in terms of size, bitrate and cycle time. To analyze these parameters on the proposed system, iMux multiplexer was used to generate a single carousel with a bitrate of 790.94 Kbps. Software tools described on section 4.1 were used to compute the values shown in Table 1. A screenshot from an application developed and running over the proposed system is shown in figure 2.

For the reference system, parameter values were measured on a frame of 3 minutes of duration, which was captured from the broadcast channel on 2008, October 24, at 9:34. It is important to remember that, in this case, each application is carried on a different carousel so each application has a different cycle time.



Figure 2. Application running on the proposed system

4.4 Systems comparison

To achieve an adequate comparison between both platforms, time measures were accomplished in two different terms. First, we measured the maximum interval of time in which an application is started from the launcher application, when the decoder has not been able to cache any files from the carousel. Second, we measured the minimum interval of time in which an application is

started once the launcher application is loaded. In this case, the decoder has been able to use its caching mechanisms to optimize the loading time. The measures obtained are shown in Table 2.

Table 1. Applications parameters

	App	Size (Kb)	Bitrate (Kb)	Cycle time (s)
Proposed system	Browser	339	218.02	12.74
	Sports	87	55.99	12.74
	Weather	166.9	107.35	12.74
	Traffic	78	50.19	12.74
	EPG	71	45.64	12.74
	Chemists	56.1	36.11	12.74
	Games	77.8	50.04	12.74
	Puzzle	112.4	72.31	12.74
	Sudoku	73.3	47.15	12.74
	Launcher	10.6	6.79	12.74
	News	93.6	60.19	12.74
	Services	64	41.16	12.74
		Total	Total	Mean
		1229.7	790.94	12.74
Reference system	Teletext	323.6	148.48	17.85
	Weather	140	98.5	11.64
	EPG	343.2	98.5	28.55
	Launcher	213.7	198.48	8.82
	Traffic	416.3	98.49	34.62
	EmpleaT	473	148.49	26.1
		Total	Total	Mean
		1909.8	790.94	21.26

Table 2. Maximum and minimum start time for comparable applications in both platforms

	App	Proposed system (s)	Reference system (s)	Difference (s)
Maximum application start time	Launcher	30	22	8
	EPG	33	97	-64
	Weather	41	56	-15
	Traffic	35	176	-141
Minimum application start time	EPG	2	3	-1
	Weather	10	29	-19
	Traffic	2	89	-87

As shown on table 2, every application on the proposed system presents a lower loading time than the same application running on the reference system, with the exception of the launcher application. The reason is that the loading time for the launcher application includes the loading time for the browser application, however such a delay will not occur anymore when loading the other applications. Furthermore, the bitrate assigned to the launcher application on the reference system is much larger than the bitrate assigned to the others. Thus, the loading time for the launcher application is improved while the loading time for the other applications get worse.

If we consider the loading time of the launcher application together with another application, the loading time for any of the applications on the proposed system is much lower than any of the applications in the reference platform. Once the launcher application is loaded, the proposed platform applications have a small and acceptable loading time. The proposed system permits the delivery of a greater number of applications on the carousel while avoiding a global bitrate increase and maintaining a proper loading time. Furthermore, although the proposed system delivers 11 applications instead of 6 applications for the reference system, the total size for the applications is reduced in 680 Kb.

5. CONCLUSIONS

In this paper, a comparative based on time performance between the proposed system and a reference system has been presented. The proposed system delivers better performance indicators, achieving better time responses while minimizing the applications size. Therefore, more interactive applications can be offered while using the same bitrate and bandwidth.

Apart of performance results, the benefits of this framework for new kinds of interactive TV services has been exposed, making it possible to use the TV as an interface to the World Wide Web.

This framework also presents benefits in development costs of interactive TV services because the complexity is reduced due to software abstraction based on XML description.

6. REFERENCES

- [1] Cineca, <http://www.cineca.tv>
- [2] DAVIC 1.4 Specification Part 9. 1998.
- [3] Document A108r1 - GEM 1.2.1 (Including IPTV). 2008.
- [4] DVB-MHP Official website, <http://www.mhp.org>
- [5] HAVi SPECIFICATION Version 1.1. May 2001.
- [6] ISO Standard 13818-6, Generic Coding of Moving Pictures and Associated Audio Information: Extension for Digital Storage Media Command and Control. 1996.
- [7] MIDP: SATSA-APDU API Developer's Guide. 2007.
- [8] MIT-xperts GmbH, <http://www.mit-xperts.com>
- [9] Morris S., Smith-Chaigneau A. 2005. Interactive TV Standards. Focal Press, Elsevier.
- [10] Multimedia and Hypermedia information coding Expert Group, <http://www.mheg.org>
- [11] NanoXML parser, <http://sourceforge.net/projects/nanoxml>
- [12] Nionex, <http://www.nionex.de>
- [13] Open Cable Applications Platform, <http://www.opencable.com>
- [14] Osmosys SA, <http://www.osmosys.tv>
- [15] Streamguru MPEG Analyzer, <http://www.streamguru.de>
- [16] Strong Digital TV, <http://www.strongsat.com>
- [17] Sun Microsystems, <http://www.sun.com>
- [18] MHP Knowledge Project. 2006. Guidelines on migration.
- [19] Tmira Solutions, <http://www.tmira.com>
- [20] TSReader TS Analyzer, <http://www.coolstf.com/tsreader>

Non Conditional Smart Card Technology to Deliver Personalized iTV Services

Giorgio Rascioni

Università Politecnica delle Marche
D.I.B.E.T.

Via Brecce Bianche 12
60131 Ancona Italy
00390712204129

g.rascioni@univpm.it

Susanna Spinsante

Università Politecnica delle Marche
D.I.B.E.T.

Via Brecce Bianche 12
60131 Ancona, Italy
00390712204894

s.spinsante@univpm.it

Ennio Gambi

Università Politecnica delle Marche
D.I.B.E.T.

Via Brecce Bianche 12
60131 Ancona, Italy
00390712204845

e.gambi@univpm.it

ABSTRACT

Among the most important innovations brought by the introduction of Digital Television, interactive MHP (Multimedia Home Platform) applications have a prominent role. This paper describes in detail how non conditional access smart cards may be used to allow the realization of personalized iTV services through properly designed MHP applications, to widen the benefits that separately come from a broadcast pervasive technology and smart card security, and thus creating synergies between information, technologies and processes. We also assess the performance of available cryptographic procedures by comparing their execution on enabled commercial digital terrestrial television receivers.

General Terms

Algorithms, Performance, Design, Experimentation, Security.

Keywords

Interactive TV, smart card, cryptography, usability, T-government, security.

1. INTRODUCTION

We live in the digital era, where the transition from analog to digital gave a strong boost to the innovations that are changing the way of creation, consumption and redistribution of services, under new business models, and through new telecommunication technologies. In this context, firms, institutions and, more in general, every type of public or private entity would like to offer advanced services to accomplish the needs of citizens, in a pervasive way. To this aim, services are requested to provide a level of personalization that implies critical issues, due to the private nature of the information exchanged. Avoiding unwanted access to personal information is fundamental to ensure a certain level of security; as a consequence, modern systems have to face issues that completely belong to the field of cryptographic technologies [1],[2].

As a general term, cryptography defines a set of techniques, procedures and algorithms that, among several possible options, allow to hide data of different nature, such as audio, video, textual data and so on, to unauthorized users. Cryptography provides efficient and powerful ways to ensure information integrity and secrecy, but the robustness of cryptographic techniques may be weakened by the difficulty to adequately protect private keys. Microprocessor smart card technology may overcome this issue thanks to the intrinsic anti-tampering protection provided in the internal microchip, mainly due to the presence of a cryptographic

co-processor able to perform the main functionalities on board [3], without releasing any key material stored in the card. As a consequence, at present the smart card technology is becoming the most viable and solid solution for the development of personalized secure services, supporting digital signature as a tool for authentication, integrity, and non repudiation, and the private key-digital certificate pair for strong authentication and confidentiality. Hence, progressively, a number of entities, especially public institutions, are migrating their own set of services into on-line, Internet-enabled service infrastructures based on the PKI (Public Key Infrastructure) architecture. A general view is shown in Figure 1, where user access to a service is provided by the release of a suited smart card.

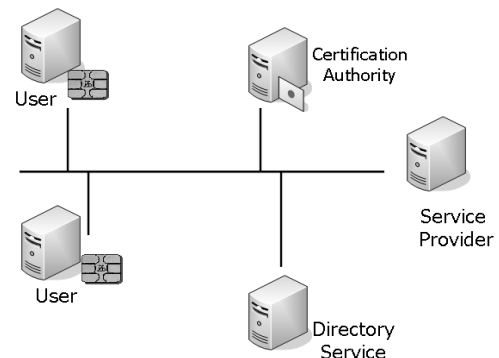


Figure 1. General view of a PKI based service architecture

In the digital TV environment, the smart card technology has been adopted since a lot of time, but only as a tool for implementing Conditional Access to specific, pay-per-view contents. Actually, a new set of applications using smart cards in the TV environment may be designed, exploiting the interactivity supported by the digital television, to widen the user experience, by simplifying the way of accessing information recovered also directly from the Internet network. Hence, the possibility to manage a non conditional access smart card by the iTV decoder opens the way to new scenarios, to move services of the Internet PKI infrastructure inside the TV environment. Several interactive services may be conceived, exploiting the peculiarities of smart card based security and authentication, such as job finding ser, learning and education , banking and financial services. The basic requirement is the possibility of accessing the smart card device through the MHP interface, in order to ensure user authentication

and protection of personal data exchanged over the available return channel.

In this paper we focus our attention, as a case study, on the possibility of extending to the iTV context services involving the use of the recent Italian smart card standard called NSC, i.e. National Service Card [4], born to favor a convergence between different digital identification solutions. Actually, the most part of our considerations is valid also for other types of non conditional smart card based solutions. In particular, we discuss possible cryptographic procedures, and their application, when the use of the NCS is extended to the interactive television context, by means of suitably developed MHP interfaces, and provide some performance evaluations focusing on user experience and comfort.

The paper is organized as follows: Section 2 briefly presents details about the MHP technological infrastructure, discussing how it is possible to manage smart cards through MHP applications, to implement the procedures made available by the NSC. Section 3 outlines the NSC chosen as a case study, its file-system design and related cryptographic procedures supported. Section 4 reports the results of experimental tests; finally, Section 5 concludes the paper.

2. SMART CARD MANAGEMENT IN MHP MIDDLEWARE

The Multimedia Home Platform [5] is a set of Java based open middleware specifications designed to add interactivity to the DVB-T transmission technology. In a digital terrestrial television receiver, the MHP middleware can be thought of as an operating system, that has the role of managing the hardware and software resources, handling and executing all the operations related to iTV applications and user interaction. Among the resources handled by the MHP middleware, the smart card reader is included, usually involved in the decoding of received pay-per-view programs within a CAM (Conditional Access Module) component, which also comprises a specific decoder for the received encrypted signal.

Almost all the commercial receivers provide at least one smart card reader slot, that is compliant to the physical and electrical standards defined in [6], and can be consequently used to communicate with a compliant smart card. Since the 1.0.3 profile, the MHP specification has introduced the SATSA (Security and Trust Services) Application Program Interface [7] as the reference environment for smart card interfacing. The SATSA package contains a set of programming interfaces dedicated to security and cryptographic functions. Among all the functionalities provided by SATSA, two sets of functions have to be necessarily supported by an interactive decoder: the SATSA Generic Connection Framework (GCF), and the SATSA-APDU package. An Application Protocol Data Unit (APDU) is the communication format between the card and the off-card applications. The format of the APDU is defined ISO specification 7816-4; a SATSA-APDU enables an MHP application to exchange APDU messages (commands or responses) with a card, according to [8]. By this way, the connection to the smart card through the GCF, and the support to exchange APDUs with the card are ensured, with the further advantage that the complex low level communication between the card and the application is made completely transparent to application developers and users. In particular, the Generic Connection Framework has been designed to supply a

homogeneous interface for different types of data connections and communication protocols.

In the context of interest, by using the GCF, an Xlet may request a connection with the card through a *factory* Java class, the *javax.microedition.io.Connector*; if the connection is established, the *Connector* returns the Xlet an object able to exchange APDUs with the related data over the established connection. The *exchangeAPDU()* method is used to send a command to a card application and receive a response: a byte array containing a command APDU is passed to the method, the command is sent to the card, and when the card sends its response APDU, this method returns the response in the form of another byte array.

3. THE NATIONAL SERVICE CARD STANDARD

The migration towards personalized services, accessible through different platforms and technologies, requires access procedures that should be secure, easy to use, and as much general purpose as possible. In Italy, to accomplish and favor the convergence among different digital identification solutions, a standard called "National Services Card" has been defined. It acts as a reference paradigm, by prescribing the minimum set of functional specifications required to a smart card for compliance, without dealing with the implementation details, nor imposing a specific technology, but allowing also the use of proprietary platforms, such as JavaCard or Multos platform. The NSC device is a microprocessor smart card based on the ISO 7816 standard, and released through a specific emission circuit, according to a standardized procedure, as prescribed in a PKI infrastructure. The NSC standard defines the file system architecture of the smart card, which is shown in Figure 2, allowing the insertion of cryptographic information necessary for authentication and digital signature operations, but also leaving free space to store information useful for supplementary types of services (bank, postal services and others).

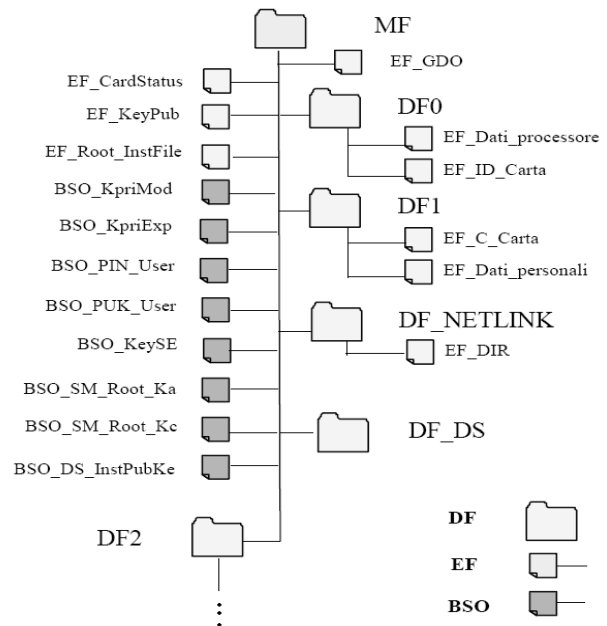


Figure 2. File system architecture according to NSC

The file system, composed by a set of Dedicated Files (DF), corresponding to directories, with the related Elementary Files (EF), is conceived in a flexible way, to allow several cryptographic operations. In particular, we can locate an area reserved to card management (including DF0, Personal Identification Number PIN, and PUK), an area to supply digital signature functions (DF Digital Signature), an area with the information necessary for authentication (Kpri, C_Card and Personal Data), and finally an area for supplementary services, under DF2. Referring to possible iTV scenarios, as outlined in the previous section, the most important basic cryptographic operations supported by the card, and available to an MHP application, are:

- *Get ATR* (Answer To Reset): this operation is necessary to reset the smart card environment, in order to open a new session with the card, for data exchange;
- *Get Data Holder*: operation that allows reading the Elementary File containing data about the smart card holder;
- *PIN Verification*: the PIN is the TEST BSO (Base Security Object) associated to binary security conditions in the card, that can be either VERIFIED (True) or UNVERIFIED (False);
- *GetCardType*: this operation performs parsing of the ATR answer to recognize the type of card inserted into the reader;
- *Get Public Certificate*: a certificate is an electronic proof which links the Signature Verification Data to a person, and confirms the identity of that person. Recovering the certificate is important because the national Italian project states that the credentials necessary to perform strong authentication are those stored in the subject section of the Public Certificate (EF-C_Card);
- *Data Signature* and *Signature Check*: operations that are equivalent to traditional handwritten signature, to subscribe digital documents.

Some of the procedures listed above are atomic, in the sense that each of them involves only one operation, or, equivalently, a single APDU command. Others are complex procedures, composed by the execution of a chain of atomic procedures. As an example of atomic procedure, we can cite the ATR (Answer To Reset) command, usually invoked to begin a new data exchange session with the card. Among the complex procedures, we can consider the digital signature functionality. A digital signature is an asymmetric cryptographic transformation of data that allows the recipient to prove the origin and integrity of the received data; digital signature protects data against forgery by third parties, and the sender against forgery by the recipient. Digital signature implementation is carried out through several transformations involving a number of APDU commands, as described in the flow-chart of Figure 3.

The data signature operation is quite complex as it results from a combination of the RSA public-key cryptosystem with the SHA1 hashing function, that provides an output of 160 bit (digest). The length of the digest does not depend on the input message length, as usual for hashing algorithms. As evident in the flow diagram of Figure 3, not all the operations are performed by the smart card cryptographic co-processor that supports only a limited number of cryptographic functions; for example Hash Computation and ID Hash encapsulation are left to the Java environment of the DTT receiver. The same happens also for *GetPublicCertificate* and

GetDataHolder operations calling the API from the *java.security* package, and performing also other operations useful for data parsing.

4. EXPERIMENTAL TESTS AND RESULTS

Compared to traditional software and, in particular, to web or desktop applications, iTV applications have different, specific requirements that should be taken into account. Most of them are due to the features of the operating systems the digital terrestrial television receivers are equipped with. They strongly affect the way an interactive application may access specific resources within the receiver, such as a modem or a smart card reader. Besides that, interactive applications designed for the iTV context shall respect specific requirements about usability and comfort of the final user: displaying information on a TV screen requires suitable design rules to be agreed upon.

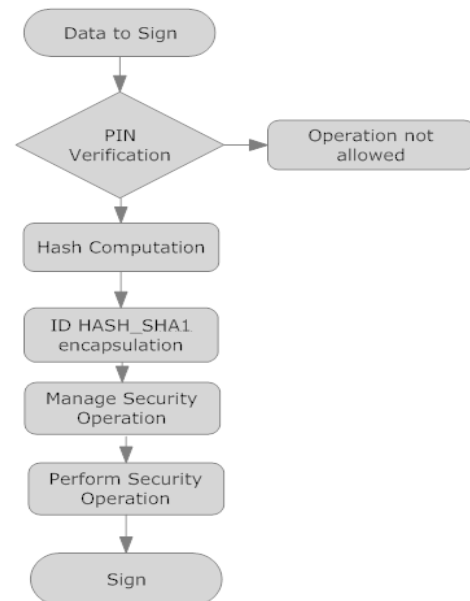


Figure 3. Flow-diagram of the signature operation

Usability of the applications can be accomplished by the definition of a set of rules about the application context and the way contents are displayed and accessed, the structure of the application, navigation strategies through the remote control, orientation, graphics, tools for advanced interaction (forms and return channel management) and, finally, some suggestions about how to write textual contents for the television [9]. The comfort felt by a user in enjoying an interactive application is mostly related to the receiver ability of providing a prompt and quick response to the user's inputs. If a receiver does not react promptly to the user's input, he will not feel comfortable and will probably switch to a different channel or content. As a consequence, interactive applications aiming at providing personalized services should be secure but also able to quickly react to the user's inputs, and to perform smart-card related operations in the shortest time. According to these considerations, we will evaluate commercial digital terrestrial receivers' performance with respect to the execution time required to perform actions related to use of a smart-card.

Experimental evaluations are performed through the execution of various routines of an iTV application written in the Java-MHP language, and adopting the SATSA API to interact with a smart card compliant to the Italian NSC standard. The routines are executed by different commercial DTT receivers, in order to verify latency or malfunctions related to the increased elaboration time required by the execution of security operations, through the interaction of the receiver with the smart card. The experimental environment available to test the way smart card based operations affect the set top box functionalities is constituted by the following elements: a smart card compliant to the Italian NSC standard, equipped with a CISC processor working at 5 MHz, three commercial digital terrestrial television receivers and a software Java-MHP application implementing the SATSA API. The first two decoders, although produced by different vendors, are based on the STi5100, mounted on different motherboards, with an amount of memory of 32 MB and with distinct firmware implementations released in different time. The third decoder is based on the STi7100, a new generation set-top box decoder chip with enhanced performances working on a motherboard with 128 MB RAM over a Linux operating system.

Table I reports the smart-card related operations that have been tested for execution by the available DTT receivers.

Table I. Execution Times

Cryptographic operation	Decoder 1	Decoder 2	Decoder 3
ATR	1 sec	0 sec	0 sec
Get Data Holder	20 sec	3 sec	1 sec
PIN Verification	3,5 sec	1 sec	0 sec
Get Public Certificate	43 sec	7 sec	3 sec
Data Signature	19 sec	3 sec	1 sec
Signature Check	44 sec	7 sec	3 sec

As shown, we tested either atomic and complex procedures, according to the description provided in Section 3. Evaluation of the execution times has been averaged over 10 experiments for each operation, and the corresponding values are reported in Table II. Results show a very different behaviour among the three decoders; these differences are not due to a single reason, but have to be addressed to the overall architecture, that comprises java middleware implementation, firmware implementation, hardware components and so on. In particular, the most recent Decoder 3, equipped with a modern chip and an high amount of memory, shows very good performances, achieving execution times near to that obtainable by a small personal computer. From a usability point of view execution times of the Decoder 1 are too long, considering that TV' users are not so tolerant. Hence, to maintain a high level general user experience of such a system, execution times have to be lower. We can assert that the execution times of the Decoder 2, with obvious reference to complex cryptographic operations that are at the base of a PKI infrastructure, represents a border line between acceptable and unacceptable performances: a TV user cannot wait more than 8-10 seconds for a procedure to

complete. On the contrary, as mentioned before, Decoder 3 obtains instantaneous reaction, thus creating the foundation for an efficient system on which it is possible to build a solid PKI infrastructure involving TV platform.

5. CONCLUSION

This paper discussed the possibility of implementing personalized iTV services on the Digital Terrestrial Television platform by the integration of smart-card based security operations with the MHP environment, thanks to suitable software interfaces provided by the Java SATSA APIs. We tested usability and feasibility of such personalized services on different platforms, to show how the execution times may vary, and influence the final perception by the user. As a matter of fact, the receiver quick reaction to user inputs is essential in ensuring a comfortable approach to iTV services, and in providing the user a sense of effectiveness of the solutions proposed. By analysing the results achieved, we can conclude that although security and usability might be in contrast, efficient software and hardware implementation of decoder elements can ensure low execution times, that means more services with a high user experience.

6. REFERENCES

- [1] Nie J., Hu X.2008. Mobile Banking Information Security and Protection Methods. International Conference on Computer Science and software Engineering ,2008 Volume3, 12-14 Dec 2008 Page(s): 587-590
- [2] Noponen, S.; Karppinen, K. 2008. Information Security of Remote File Transfers with Mobile Devices. 32nd Annual IEEE International Computer Software and Applications, 2008. COMPSAC '08. July 28 2008-Aug. 1 2008
- [3] Cooke, J.C.; Brewster, R.L. 1993. The use of smart cards in personal communication systems security. Fourth IEEE Conference on Telecommunications, 1993. 18-21 Apr 1993 Page(s): 246 – 251
- [4] CNIPA Centro Nazionale per Informatica nella Pubblica Amministrazione (in Italian). DOI=<http://www.progettocns.it/cittadino/usaCarta.aspx>
- [5] ETSI ES 201 812 V1.1.2 “Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.0.3”, August 2006.
- [6] International Standard Organization, “Smart Card Standard 7816-1: Identification cards - Integrated circuit(s) cards with contacts - Part 1: Physical characteristics”, 1998.
- [7] “Security and trust services API for J2ME (SATSA); JSR177” [Online]. DOI= <http://java.sun.com/products/satsa>.
- [8] International Standard Organization, “Smart Card Standard 7816-4: Identification cards - Integrated circuit(s) cards with contacts - Part 4: Interindustry Commands for Interchange”, 2005.
- [9] Fondazione Ugo Bordoni-FUB (2005). “*Raccomandazioni per le interfacce dei servizi interattivi della televisione digitale*”, Working Group GLAD1. DOI= http://www.iconmedialab.it/pdf_interattive/interfacce_interattive.pdf

PersonalTVware: A Proposal of Architecture to Support the Context-aware Personalized Recommendation of TV Programs

Fábio Santos da Silva
Polytechnic School of

University of São Paulo (USP)
AV. Prof. Luciano Gualberto,
travessa 3,n158, sala C1-46,
05508-900, São Paulo, SP, Brazil

fsilva@larc.usp.br

Luiz Gustavo Pacola Alves
Polytechnic School of

University of São Paulo (USP)
AV. Prof. Luciano Gualberto,
travessa 3, n158, sala C1-46,
05508-900, São Paulo, SP, Brazil

luizgpa@larc.usp.br

Graça Bressan
Polytechnic School of

University of São Paulo (USP)
AV. Prof. Luciano Gualberto,
travessa 3, n158, sala C1-46,
05508-900, São Paulo, SP, Brazil

gbressan@larc.usp.br

ABSTRACT

The arrival of Digital TV has ensued the growth in the volume of TV programs offered by TV operators increasing the difficulty by users in locating relevant content. In addition, TV users are not required to have as their main task to search for information as it is the case of the Internet. Within this scenario, the recommender systems stand out as a possible solution for this problem. However, the context has rarely been explored during the recommendation process. This paper presents a proposal of architecture of support to context-aware personalized recommendation for Digital TV – entitled PersonalTVware. Finally, a recommender process is described grounded on the proposed architecture.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Information filtering, Selection process*;
H.4 [Information Systems Applications]: Miscellaneous.

General Terms

Algorithms, Design, Experimentation

Keywords

Recommender Systems, Context-Awareness, Interactive Digital TV

1. INTRODUCTION

One of the consequences of the digital revolution was exactly the increase in content options offered through Digital TV (terrestrial, satellite and cable) [13]. Thus, the difficulty in finding the content the user wishes in a transparent way among the available TV program option increased [3]. The traditional tool known as Electronic Program Guide (EPG), therefore, has not efficiently responded to the needs the user has for information. EPG simply displays long lists of TV programs requiring the user to spend a great deal of time looking for information on his favorite programs. Due to the large number of content, finding information on TV programs via EPG by means of a remote control has turned into a hard task. In view of the above-mentioned problem related to the information overload

vis-à-vis the Digital TV [2], the execution of new studies for the development of systems whose purpose is to enable the access to TV programs through a personalized approach is justified. Thus, several works have proposed solutions based on the classical approach of content personalized recommendation [2] [3] [13].

However, most recommender systems for Digital TV have rarely considered the user's context information in carrying out the recommendation [1], which restricts the quality of the generated results. Usually, the recommendation process is based on the level of interest either explicitly obtained or implicitly inferred from the use background of individual or the community of people having similar interests. Without the notion of context, the level of interest can only provide general recommendation.

Thus, this work proposes that some questions related to context can be exploited such as, **who** is the user who is watching television at that moment? **Where** is the user located? In his house? In his office? **How** is the user watching television? Through a mobile, portable or fixed device? **When** does he usually watch a specific type of TV program? In the evening upon arriving at home from work or on Sunday morning? **What** content is considered relevant by the user? Depending on his context, the user may have different preferences and needs.

The research object consists in the investigation of how such contextual dimensions can be exploited during the personalized recommendation process. This work, therefore, is based on the hypothesis that the context can influence the user's preferences when he watches television and proposes the exploration thereof as a promising approach aiming at improving the quality of recommendation in the Digital TV scenario. This article presents the PersonalTVware architecture in order to enable the user to receive the personalized recommendation of content at the right and adequate time and place to the access device.

2. RELATED WORK

PTV [3] was one of the pioneering personalized recommendation projects for Digital TV content, and it has become a reference landmark for many initiatives, which investigate Digital TV information overload. PTV is a system providing an Internet-based personalized TV listings service. Differently from

PersonalTVware designed for both middlewares of European and Brazilian standards – DVB-MHP [9] and GINGA [8], respectively. Zhang and Zheng [13] have developed TV programs personalized recommendation based on the specification of TV-anytime standard metadata and on the Content-based Filtering technique. A fuzzy inference system was used as a method for automatically analyzing the user's preferences obtained from the usage history. Blanco [2] has submitted a proposal for a personalized content recommendation approach that exploits the Semantic Web technologies. This new approach was proposed in the AVATAR project, within which a personalized recommendation system that makes use of MHP, TV-Anytime and Web Ontology Language (OWL) was developed.

In contrast with the PersonalTVware, the above-mentioned works fail to exploit the user's context during the recommendation process. Thus, the investigation of the intersection of the recommendation systems as the context-aware systems for the provision of relevant content has increasingly received more attention from the academic community [1].

3. PERSONALTVWARE PLATFORM

This work has been based on the hypothesis that the user may have different preferences as to content depending on his current and also past context. There is a classic definition of context given by Dey [4] that is most accepted and referenced in the academic community: "Context is any information that can be used to characterize the situation of entities. An entity, by itself, can be a person, place or object that is considered relevant for the interaction between a user and an application, including the user and the application itself". As such, in the Digital TV scenario, to be aware of the context information of the user's interaction is relevant for the personalized selection and adaptation of contents to be presented to the user [5]. The user's context can be managed in order to determine what type of TV program is more adequate for recommendation.

3.1 Example Use Scenario

To illustrate the applicability of the PersonalTVware, a use scenario will be initially presented.

"Soon as she gets to the Gym on Monday at 07:30 P.M., Fernanda begins her treadmill work out. She enjoys watching TV programs on sports in her portable device while she works out. Nevertheless, Fernanda does not wish to waste time looking for a TV program through various TV channels. She, therefore, uses to access information on the Digital TV schedule in a personal way by means of a recommendation system that presents a list of recommended TV programs in accordance with her profile and current context."

3.2 Context Identification

Through the presented use scenario it is possible to identify some context information implicitly present such as "Fernanda", "Monday at 07:30 P.M.", "gym", "treadmill work out", "portable device". Such information has to do with contextual dimensions: **who** (identity), **when** (time), **where** (location), **what** (activity) and **how** (a way to identify how the elements of context are collected). To efficiently exploit the user's context, a context

model geared towards the organization of the main context information extracted from the use scenario was specified. The model is used as a reference in the construction of metadata structures in XML Schemes [12] used in the representation of context information.

3.3 Architecture

The architecture proposed was designed to offer support to the development of applications of context-aware personalized recommendation in a modular and flexible way. The architecture comprises two subsystems: the user's device and the service provider. The user's device subsystem could be implemented in a set-top box, a portable computer or a mobile phone, or else, a Mobile TV equipped with an onboard middleware such as GINGA or MHP. The communication between the subsystems will be bidirectional in format through a service interface based in a Web service using HTTP and TCP protocols under the return channel of the device to transmit and receive metadata. Figure 1 illustrate the architecture with their respective modules.

In the user's device, the Recommendation Manager module interfaces between the client applications and the other modules of the architecture, and it is responsible for the management of the recommendation process by coordinating the other modules of the system. When the user wishes to be informed as to which TV programs have been recommended, he is required by means of a request by the application software to the manager in order that the context-aware personalized recommendation process as described under subsection 3.4 be executed. In addition, the Relevance Feedback [2] also falls under the action of this module enabling the user to assess from among the recommended TV programs those it considers either relevant or irrelevant. This way, this functionality will allow for the expansion the choice restrictions imposed by the filter refining future recommendations.

The User Context Manager module is responsible for the access, acquisition in an implicit way and conversions of the information within the user current context, fact that should generate a history of past contexts. The user-context information will be represented by means of metadata structures in XML Schemas based on the specified-context model.

The User Profile Manager module is responsible for the access and acquisition of the information that constitute the user profile in an explicitly way. Through its components the user will be able to specify information such as personal data (name, age, gender, occupation) and preferences (TV programs, director, actor, topics, among others). This information will be described in accordance with the specifications of metadata of the TV-Anytime [10] and MPEG-7 [6] standards making the representation standardized and structured. In addition to this, owing to privacy and safety reasons, the user profile will be stored only in the user's device.

The User Context Interpreter module is responsible for inferring implicit preferences per channels and TV program genres from the interpretation of the current and past context information obtained through User Context Manager module. The approach used for performing the inference is based on the use of a set of rules.

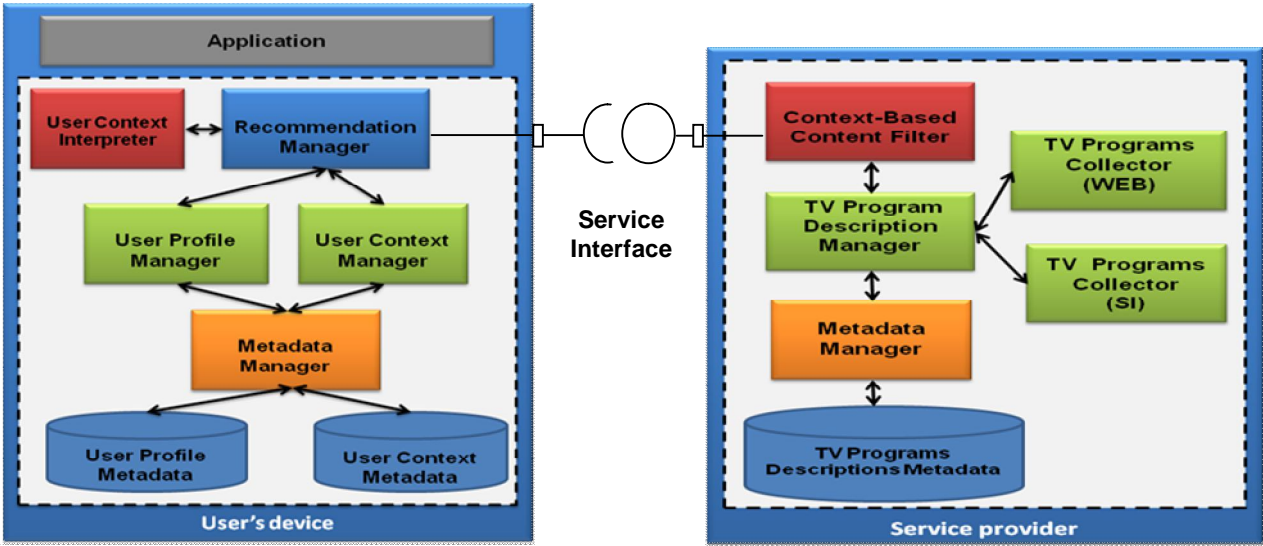


Figure 1. PersonalTVware Architecture

Upon the study of several inferencing methods, the rule-based reasoning method [11] has been selected, since it allows for the definition, in a flexible way, of conditional structures, which reflect the relations among contextual dimensions. If a determined user-context is true, then, corresponding TV programs genres and channels are inferred. For example, if the user's identity 001, recommendation requests on Sunday at 08:00 P.M. in his apartment, using a fixed device (set-top box) then he should have an interest in TV programs of the kind of movie-comedy and HDTV channels.

It is also possible to have more than one genre associated to a context condition. This way, asset of rules or contextual preferences can be explicitly defined by the user in accordance with the context variables: location, day, time of interaction and type of access terminal. The inference task for implicit preferences from the history of past contexts is possible through case-based reasoning (CBR) [7], where the context variables (location, day, time of interaction and type of access terminal) define a case. In this type of technique the user preference for a new context is inferred grounded on similar contexts (cases), which had occurred in the past. The context information is captured in an implicit way through the User Context Manager module.

Because of the limitation of the computing resources of the access devices, the Context-Based Content Filter module is located in the services provider. This module is responsible for the filtering of TV programs, which will likely be relevant for the user considering his current context. This way, the filtering process exploits contextual variables (day, time), user profile, his inferred implicit preferences and the descriptions of the contents of the TV programs. The used information filtering has as its base the Content-Based Filtering technique. The TV Program Description Manager module is responsible for the consultation and insertion of information relative to the TV programs. Such information is also described in accordance with the TV-Anytime standard metadata specifications. In both subsystems: user device and services provider, the Metadata Manager module provides the support for the other modules in the architecture, and is

responsible for retrieval, storage and validation of the metadata represented in XML documents. The XML format has been adopted because of the integration with the metadata standards used in the PersonalTVware. This module, therefore, acts as a mediator between the other modules and the database by supplying a set of methods, which enable the handling of metadata in a transparent way.

Finally, the modules: TV Programs Collector (WEB and SI) should be used for capturing information relative to TV programs from outside sources such as WEB and SI (*Service Information*). SI is organized in the shape of tables, which include metadata on services, events, date and time, duration, among other. Thus, the WEB collector allows the system administrator, through a WEB interface, to submit the metadata relative to TV programs. The automatic capture of the schedule grid published in the sites of broadcasters is an alternative that is being investigated. The SI collector is responsible for extracting the metadata from the SI tables provided by each TV broadcaster from a TV signal receiver unit linked to the system. This way both modules are responsible for the automatic updating of the TV program metadata base by means of the submission of the extracted metadata to the TV Program Description Manager module.

3.4 Recommendation Process

Figure 2 illustrates the context-based personalized recommendation process through a diagram of activities. Assuming that the user had explicitly defined his profile, the process begins when the Recommendation Manager module receives a request to retrieve a list of TV programs which have been filtered in accordance with the user profile and implicit preferences inferred from user context identified through his ID. The User Context Manager module is activated to capture and represent the user's current context information such as user's identification, location, day and time of the interaction and type of access device, which should generate a history of the user's contexts. Later, the Recommendation Manager module, having the user's ID checks its respective information from the current context and the user's profile via the User Context Manager and

the User profile Manager modules. Such context information is forwarded to the User Context Interpreter module that should infer implicit preferences by TV program genres by means of a rule-based reasoning method so that the user may receive recommendations as to TV programs, which are adequate to his current context. The Recommendation Manager module receives from the User Context Interpreter the implicit inferred preferences and, then, joins the explicit preferences and personal information as defined in the user profile and sends them to the filtering module. Finally, the Context-based Content Filter module carries out the filtering of the TV programs through the comparison between the contextual variables (day, time), user profile, and his inferred preferences along with the descriptions of the TV programs obtained through the TV Program Description Manager module. This way, upon receiving a request featuring the profile and implicit preferences inferred from a specific user as input, the module executes the filtering of TV programs in order to obtain a list in decreasing order for TV programs based on the value of similarity which will be later forwarded to the Recommendation Manager module.

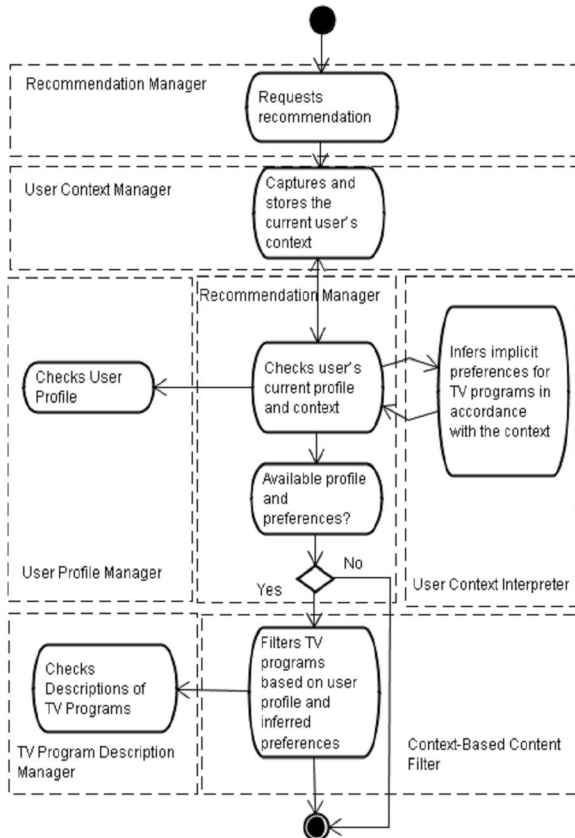


Figure 2. Recommendation Process

4. CONCLUSION AND FUTURE WORK

This paper has attempted to highlight the possibility of exploiting the user's context during the personalized recommendation process as an approach to deal with the information overload problem within the Digital TV scenario. A general view of the proposed architecture for the PersonalTVware was presented

with the objective is to provide a support to the development of application software for context-aware personalized recommendations. A recommendation process that exploits the architecture was further presented. This work, therefore, ought to contribute to the Digital TV interactive application development area that has been increasingly calling the attention of the world academic community, particularly that in Brazil. To the extent digital TV takes hold in Brazil, demand and supply for interactive services will occur, within which personalization will play a major role. Both personalized recommendation application prototypes will be developed in future works and experiments taking into consideration a real environment comprising several interacting users will also be carried out featuring the personalized recommendation applications within several contexts.

5. REFERENCES

- [1] Baltrunas, L. Exploiting contextual information in recommender systems. In: Proc. of the ACM conference on Recommender systems, pp 295-298, 2008.
- [2] Blanco, F. Y., et al. AVATAR: An Improved Solution for Personalized TV based on Semantic Inference. IEEE Transactions on Consumer Electronics, vol. 52, no. 1, pp 223-231 - feb 2006.
- [3] Cotter, P. e Smyth B. PTV: Intelligent Personalized TV Guides. In: Proceedings of the 12th Innovative Applications of Artificial Intelligence (IAAI-2000) Conference. AAAI Press.
- [4] Dey, A. K. Understanding an Using Context. In: ACM Personal and Ubiqu. Comp. Journal, vol.5, no.1, pp.4-7, 2001.
- [5] Goularte, R. Context-based content personalization and adaptation for Interactive TV. Phd thesis. Institute of Mathematics and Computing, University of São Paulo, São Carlos, 2004.
- [6] ISO/IEC 15948-1 MPEG-7 Overview, 2004.
- [7] Lorenzi, F.; RicciI, F. Case-Based Recommender Systems. In: John Wang. (Org.). The Encyclopedia of Data Warehousing and Mining. I ed. Hershey: Idea Group Reference, pp. 124-128, 2005.
- [8] Souza F. G., et al. Ginga-J: The Procedural Middleware for the Brazilian Digital TV System, Journal of the Brazilian Computer Society, vol. 12, no. 4, 2007.
- [9] Morris, S.; Smith-Chaigneau, A. Interactive TV Standards: A Guide to MHP, OCAP, and JavaTV. Focal Press, 2005.
- [10] TV-Anytime Forum. TV-Anytime – Part 3: Metadata, 2007 ETSI TS102 822-3-1 v1.4.1 Technical Specification.
- [11] Watterman, D. A. A Guide to Expert Systems. Addison Wesley, 1986.
- [12] W3C. XML Schema. W3C Recommendation, 2001 DOI: <http://www.w3.org/XML/Schema>.
- [13] Zhang, H.; Zheng, S. Personalized TV Program Recommendation based on TV-Anytime Metadata. In: Proc. of the Ninth International Symposium on Consumer Electronics, (ISCE 2005), 2005.

From Analogue to Digital: Television Sound moving towards New Media. Stationary Sound Applications and Mobile Sound Applications

Ahti Korhonen

Doctoral student: University of Jyväskylä, digital culture studies

Sound engineer: Finnish Broadcasting Company

Address: Maustetehtaankatu 9 B 22

00990 Helsinki, Finland

Tel: +358-50-411 6151

ahti.korhonen@yle.fi

ABSTRACT

The television sound is going through major changes. Analogue television sound has been broadcasted **one-way** – from broadcaster to receiver and the idea of broadcasting has been **top-down**. Now digitalization of television sound is going to change from receiving to using and adjust. The purpose of this text is to show what kind of sound applications there will be as available in future television. The digital television has good possibilities to offer several types of sound services. This paper is a part of my dissertation about television sound moving towards new media. The Stationary Sound Applications and Mobile Sound Applications are the most central issues when television sound is going from analogue to digital.

KEYWORDS: Audio, broadcasting, digital television, Hi-Fidelity, interactivity, sound quality

1. INTRODUCTION

Television sound has been analogue since the 1930s. Since then it has moved toward digitality almost 30 years after the 1970s. The digital television is taking its first steps offering more in sound.

There has been convergence in communication media during the past 20 years but not much attention has been paid to sound. Talking about Interactivity, sound is mostly connected to the picture. But PCs and digital television integration offer more opportunities to interactivity. Of course digital television is a more suitable arena for easy sound applications than PC.

Surround Sound is one application for digital television and we have already seen applications coming through (for minorities; text-to-sound) in digital television. In the future, there will a larger variety of applications. Surely broadcasting has its limits when offering sound applications.

As a whole, television sound is becoming digital. Since 1990, television programs have been produced and done digitally. They are received digitally as well. Surround Sound must be seen as a means of using the sound, in stead of just listening to it.

2. RELATED WORK

Most researches of digital television are related in visual aspects, usability, economics or broadcasting systems etc. In sound, the researches have been technological, being associated to Surround Sound, acoustical measurements and speech researches. Aesthetics, media theories' or historical points of view towards new media have rarely been subjects of research.

Some researchers have done dissertations about digital television like Pertti Näränen, Seppo Kangaspunta, Leena Eronen and Kari Jääskeläinen, but none of these is related to sound. Amitabh Kumar writes very much about broadcasting systems of digital television and he has small sections about applications to digital television and mobile phones.

About sound history there is very much material of sound broadcasting and sound itself. AES (Audio Engineering Society) got much articles about sound quality and technical data.

3. BROADCASTING SYSTEMS TO DIGITAL TELEVISION

The basic idea of broadcasting systems can be called today as a *net*. Paul Barans' thinking (Gere 2006, 68) when talking about the net has three aspects. In the first model there is a central unit and many receiving points. The second model includes many broadcasting units and the idea is non-centralised. The third model is a pure net. These aspects of broadcasting and receiving has its ideological points of view.

Today when broadcasting sound and vision we use four categories of broadcasting modulation standards: DVB-T, DVB-S, DVB-C and DVB-H. Broadband will be the next system for broadcasting the HDTV. It is said that its capacity will be 100 Mb/second or more. This huge amount of capacity has room for future thinking with applications – the sound applications have to get more room when broadcasting and interactivity has to get its capacity.

The basic idea of broadcasting is simple: TV and radio channels send digital encoded MPEG materials to the multiplex and the byte stream will continue to the antennas, the received broadcasted program then being translated back to the analogue audio signal. Byte rate reduction and compression will be used. Broadcasting and receiving the MPEG – 2 stream is coded and decoded and the multiplexed material includes the program information (Crinon,

Bhat, Catapano, Thomas, Loo & Bang 2006, 115 – 116).

Within the standardisation of broadcasting systems there are possibilities of serving many kinds of applications in digital television as well in broadband and mobile systems. Serial Digital Audio Interface standard was agreed jointly between Audio Engineering Society and European Broadcasting Union 1985 and this specification allows distribution of uncompressed stereo audio signals as a serial digital bitstream (Jones, Defilippis, Hoffman & Williams 2006, 25).

4. SURROUND SOUND AND DIGITAL TELEVISION SOUND QUALITY

Surround Sound is coming to the digital television environment. It has got its basic idea from the cinema sound.

We should talk about the Hi-Fi (High Fidelity) as a normal situation with television sound. Byte reduced sound and compression does not always offer the best alternative to the sound. In television we have got monophonic, stereophonic and sound dubbing. In Finland the Surround Sound (AC3) will get a capacity 448 kb/second. This is enough to get a good quality of sound.

A strong byte reduction combined with a low sampling rate will result in a poor sound quality. When the byte reduction is low and the sampling rate is high, the sound quality will be good. In general, the packing of sound seems to cut the top quality off from the sound – what is called Hi-Fi.

There are still other ways to determine the sound quality of digital television. Poor frequency response, unclear sound field, disturbances in broadcasting and poor channel separation.

A good sound quality will be a result of low byte rate reduction. This should apply to the whole broadcasting chain. This can be reached by using a high sampling rate for digital audio. Of course good frequency response after that and good channel separation are fields that should be taken into consideration.

Good sound quality could mean other things, such as high quality sound applications and a functional broadcasting system. Functionality is the first requirement for functional Surround Sound.

5. STATIONARY SOUND APPLICATIONS AND MOBILE SOUND APPLICATIONS

Sound Applications are divided in two categories. The first one is Stationary Sound Applications and the second one is Mobile Sound Applications.

Stationary Sound Applications are used in home environment, including PC, Internet, digital television and connections between them. Sound equipments will be connected to PC and to a Set-top box. Sound systems are fully digitalized. There will be next kind of applications to home environment: packed files, alert sounds, games,

sound services, guide tones, signal tones, music libraries, speech recognition etc.

Mobile Sound Applications mean using sound applications when going outside and to public places. All mobile systems can be connected to cellular phones and laptops. Public transportation can be equipped with wireless systems. There will be next kind of applications in public places: sound navigation (Nokia 2008), packed files, radio and tv programs, news.

Joseph Weber and Tom Newberry (2007, 86) have dealt the multimedia in following categories: video conferencing, streaming audio (audio packing), Internet telephony, streaming AV-services (IPTV). When transferring audio packages to the customer, the stationary or mobile capacities are quite small. Large amounts of bytes are not required to get the applications to your cellular phone or digital tv.

DVB – H (Digital Video Broadcasting – Handheld) is mobile broadcasting where the radio and television programs are sent and listened to in mobile phones and laptops. The interactive services are there as well. DVB – H uses 2G and 3G nets as a return channel (Cave & Nakamura 2006, 4.)

6. DIGITAL TELEVISION AS A POSSIBILITY

Richard Sennett (2005, 143) has described sound equipments as a physical possibility. iPods' commercial success is related to its' great capacity to save so much music that no-one is able to listen it through during one day. A similar allurements should be found for digital television to make it easy to use the sound applications or sound services.

Stationary Sound Applications are a part of digital television sound system in home environment. Sound applications mean interactivity with digital television or PC, if connected to digital television.

From digital television we can get television and radio programs, communication services, information services and other sound applications. When designing these solutions we have to seek interactivity and understand technology and user-interfaces. We also have to pay attention to user groups and products that we are serving.

Digital television's Sound applications could provide both public and commercial applications. These two sectors could offer services to citizens using both stationary and mobile solutions. There could be several kinds of services, such as information, archives, signals, alarms etc.

Since the beginning of year 2000, there have been efforts to get the MHP solutions to the markets and applications have been created for the needs of the banking sector, companies, marketing, games, multimedia and video conferencing etc.

Digital television systems belong to the Stationary Sound Application environment. The connection between Personal Computer and digital television gives a variety of possibilities for the user. These are CD, DVD, sound applications, sound services, information etc. Connection with PC brings multimedia available. Music, games, extra information (advising & service) can be used through digital television.

Brecht and Kunert (2005, 163) see that when talking about sound there should be at least the play, stop, rewind and forward functions. In addition to these several kinds of other functions should be available. Tuning, the equalizer, marking the starting point and the end of material be there as well.

John Carey (2003, 54) wrote that ITV can be made to increase people's own convenience rather than requiring them to follow a fixed television schedule. Flexibility will be improved when using sound applications.

7. CONCLUSIONS

The sound applications should be easy to use and user-interfaces must be simple. EPG is a suitable user-interface. Then the broadcasting capacities should be high enough with the broadband. And of course the interactivity should be found with the feedback possibility.

Before you have just watched and listened to the television – now your connection is moving towards using and adjusting.

Packing files offers the possibility to use the files once again and adjusting the files is a new possibility to find something new. Stationary Sound Applications and Mobile Sound Applications will be the next generation of going digital with television sound. Understanding, using and adjusting the sound, as an interactive methodology, makes the user active. Getting information in sound gives many kind of possibilities within both, stationary and mobile services.

8. REFERENCES

- [1] Brecht R. & Kunert T. 2005. User requirements and guidance for interactive TV news applications. Teoksessa: Euro ITV. 3rd European Conference on Interactive Television. InDiMedia.
- [2] Carey J. 2003. New media and TV Viewing Behavior: Implications For Public Broadcasting. In: NHK broadcasting studies; an international annual of broadcasting science. NHK Broadcasting Culture Research Institute.
- [3] Cave M. & Nakamura K. 2006. Digital Broadcasting. Policy and Practice in the Americas, Europe and Japan. Edward Elgar Publishing Limited. UK.
- [4] Crinon R. J. & Bhat D. & Catapano D. & Thomas G. & Loo J. T. Van & Bang G. 2006. Data Broadcasting and Interactive Television. In: Global digital television: technology & emerging services. Proceedings of the IEEE. January 2006. Vol. 94. No. 1.
- [5] Gere Charlie. 2006. Digitaalinen kulttuuri. Translation: Raine Koskimaa and working group: Jussi Parikka, Petri Saarikoski, Tanja Sihvonen, Jaakko Suominen, Juha Wakonen. Original book 2002. Faros-kustannus Oy. Turku.
- [6] Jones G. A. & Deffilippis J. M. & Hoffmann H. & Williams E. A. 2006. Digital Television Station and Network Implementation. In: Global digital television: technology & emerging services. Proceedings of the IEEE. January 2006. Vol. 94. No.1.
- [7] Nokia. 2008. Uudenajan laitteet ja palvelut. Nokia-talvikuvasto 2008-2009. www.nokia.fi
- [8] Sennett R. 2007. Uuden kapitalismin kulttuuri. Translation to Finnish: Kaisa Koskinen. Original book: The Culture of the New Capitalism. Osuuskunta Vastapaino. Tampere.
- [9] Weber Joseph & Newberry Tom. 2007. IPTV Crash Course. The McGraw-Hill Companies. USA.

Generation, Streaming and Presentation of Electronic Program Guide

Cyril Concolato

Télécom ParisTech; Institut Télécom; CNRS LTCI
46, rue Barrault, 75013 Paris, France
Cyril.Concolato@telecom-paristech.fr

ABSTRACT

With the deployment of IPTV and mobile TV systems on the one hand, and the diversity of devices capable of displaying rich-media content on the other hand, the traditional monolithic way of presenting electronic program guide (EPG) data is becoming inappropriate. In this paper, we describe a system which proposes to separate the generation of presentable EPG data from its actual display. We present a system which achieves the generation and streaming of EPG based on streamable declarative languages. This system allows a reactive distribution and the efficient presentation of dynamic EPG.

Categories and Subject Descriptors

I.3.2 [Computer Graphics]: Graphics Systems – *Distributed/network graphics*; I.7.2 [Document and Text Processing]: Document Preparation – *Hypertext/hypermedia, Standards*

General Terms

Design, Experimentation, Languages, Performance.

Keywords

Electronic Program Guide, Presentation Language, Streaming.

1. INTRODUCTION

Traditional Electronic Program Guides (EPG), displayed on today's televisions, are the result of the on-the-fly generation of an interactive presentation based on information extracted from the broadcast channel. The generation and presentation are performed in a single place using software embedded either in the TV or in a STB, developed using programmatic languages. In traditional DVB environments, the EPG information is extracted using Event Information Tables (EIT) [1]. With the deployment of IPTV and mobile TV systems, this ecosystem is currently experiencing many changes. First, the number of sources, the type and quantity of information is changing. For example, EPG information can be retrieved from Web feeds (e.g. XMLTV [8]) or from DVB-IPDC [2] channels in the form of TV Anytime data [11]. Second, the number and type of devices capable of

displaying rich media content (such as EPG) is also changing: TV, PC, PDA, Smartphones, Portable Media Players (PMP). All these devices have different screen configurations, interaction methods, connectivity features or processing power. The existing method which consists in embedding monolithic software to process and display EPG from a single source is therefore not well suited anymore because of development costs. As a consequence, more and more solutions now rely on dedicated and tailored Web browsers such as [3] to display EPG.

The generation and display of EPG is not a new research topic and many papers have been written on the topic. Most of them focus on the design of efficient interactive paradigms [14] or on recommendation systems [16]. In this work, we show a system which decouples the generation and the display of the EPG. This can be viewed as an example of the secondary screen approach described in [17]. The EPG is generated in a device, different from the rendering device, tuned to the broadcast channel to retrieve the raw EPG data. This data is then transformed into a presentation form which is streamed to the remote rendering device (phone, PMP). In this paper, we use the 3GPP DIMS [4] declarative language, rather than XHTML, to first benefit from its light, declarative rich graphics and animation features but also, to allow streaming of the EPG. The use of streaming follows the push approach of broadcasting technologies used in the delivery of raw EPG data but also allows guarantying a strong synchronization between the presentation and raw data. Additionally, streaming, coupled with the DIMS update mechanism enables light presentation processing in the client.

In the remaining of this paper, we present in Section 2 the possible architectures for such a system. Then, in Section 3, we detail our system. In Section 4, we show and discuss some results of generated EPG. Finally, we conclude this paper and propose future work in Section 5.

2. EPG GENERATION AND PRESENTATION ARCHITECTURES

This section presents existing approaches for the generation, delivery and presentation of EPG. We focused only on approaches where a declarative language is used to describe the presentation. From the literature [15] or from existing products, we propose to classify the approaches into two main extreme approaches, depicted in Figure 1. We note that this classification is actually rather generic and could be applicable to the generation, streaming and presentation of many types of metadata other than EPG. For both approaches, we assume that the metadata comes from either some broadcaster content management system or is dynamically produced, as for live events. In both cases, we also

assume that a metadata filtering mechanism can be used at the server-side or at the client-side either driven by direct user interaction (e.g. show sport programs only) or based on user modeling systems.

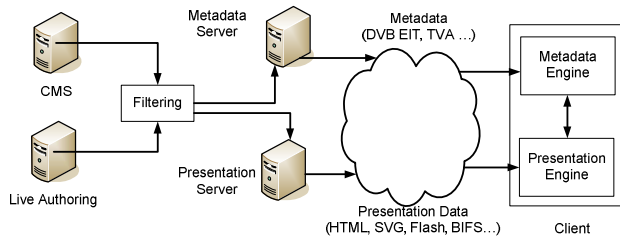


Figure 1 – Alternative approaches for the generation, delivery and presentation of EPG

2.1 The Metadata-Driven Approach

In this first approach, the EPG information is delivered in the form of metadata, and no information is sent about how this metadata should be presented. The receiver is in charge of presenting the metadata in a suitable form. This process can be driven by some specific generation software which creates a declarative presentation (e.g. using XML Transformations). This presentation is then passed to the presentation engine for display, typically a browser. Alternatively, this process can be driven directly by the presentation engine. In that case, the presentation engine loads some presentation templates, and then fills the template with data from the metadata engine. Such metadata, e.g. the name of the current program, can be pulled using some specific API (such as the Joost Widget API [7]); or if the metadata engine is implemented as a web server, by using the XMLHttpRequest standard (XHR) [6]. In an other alternative, the metadata can be pushed to the presentation engine.

The advantages of the metadata-driven approach are the following. First, since the presentation data is generated at the receiver side, it can be easily adapted to the terminal characteristics (e.g. screen size, input methods) and to the user preferences (metadata of interest, user model) without privacy concerns. Second, since this approach delivers ‘raw’ metadata without presentation data, it is efficient in terms of bandwidth. Third, from a server point-of-view, the implementation is simple since it does not have to deal with presentation data.

The short-comings of this approach on the other hand are the following. First, in order to display the presentation, the receiver must implement a metadata engine and the associated interface to the presentation engine. In the context of the growing number of sources and formats of data, it means that the receiver must implement a metadata engine capable of handling several types of metadata. An alternative would be to aggregate all the metadata in one unique form. In some sense, this alternative would transfer some intelligence from the client to the server. Second, in this extreme approach, the presentation look and feel is driven by hardcoded presentation parameters stored on the client. It means that the presentation style cannot change over time (changing colors, fonts or even navigation scheme). Third, since each receiving device can use presentation different parameters, the metadata broadcaster has no control over how the metadata will be presented on the receiving terminal. We believe this is an

important problem which leads towards using an approach where presentation data is sent to the receiver.

2.2 The Presentation-Driven Approach

In this second extreme approach, the broadcasting server aggregates the different sources of metadata, transforms them into presentation data, and finally delivers the presentation data to the presentation engine of the client. Oppositely to the previous approach, no ‘raw’ metadata is sent to the client. In this approach, the presentation data can be queried from or pushed by the server.

This approach has the following advantages. First, the client does not implement a metadata engine. Its footprint is therefore smaller. Second, with this approach, it is possible to update the presentation of the EPG for example to change the look of the EPG during the Christmas time or the Olympics, or to change the navigation method. Third, in this approach, the presentation can be displayed in the receivers as the author decided.

However, this approach also has limitations. First, it requires more processing at the server side. Second, the diversity of rendering devices hardens the task of adapting the presentation to the terminal characteristics (presentation formats, screen size, input methods). Third, since no metadata is sent, there are cases where the loss of semantics in the presentation data prevents the client from performing semantic filtering (e.g. showing only Sports) or makes it more complex. Finally, since the metadata is transformed into presentation data, a higher bandwidth is required to transmit the presentation style and navigation paradigm.

2.3 Summary

Table 1 summarizes the comparison between the previous approaches. Obviously between those approaches, many hybrid approaches can exist. One interesting approach could consist in sending some metadata along with some presentation information to control the presentation (at the cost of bandwidth occupancy).

Table 1 – Comparison between the metadata and presentation-driven approaches for the generation and presentation of EPG

	Metadata-driven	Presentation-driven
Server Complexity	less	more
Client Complexity	more	less
Bandwidth required	less	more
Adaptation handling	simple	complex
Client-side semantic filtering	simple	complex
Presentation control	no	yes

There are several options to do that. One could leverage the ability of the languages like XML to mix data. In that case, a server could deliver a mixed document with presentation data, EPG metadata and the logic to build the final presentation. Another option would be to use separate the data into files or streams and to link the presentation data to the metadata. Files could be acquired using XHR or using file delivery mechanisms such as the DVB Object Carousel [9] or the FLUTE protocol [10]. Streams could use the MPEG-7 BiM [12] standard for metadata and for the presentation data, 3GPP DIMS (MPEG-4 LAsEr) or MPEG-4 BIFS [5].

3. THE PINGO SYSTEM

Our system, called PINGO, builds upon the previous analysis. The use case for the PINGO System is the in-house redistribution of television signals, including the EPG information. Indeed, the deployment of mobile TV and terrestrial TV do not allow, for now, for correct in-door reception of the signal by devices like mobile phones or Portable Media Players. In the PINGO system, a centralized device, named PINGO Box, standalone or integrated in a Set Top Box or in an Internet Service Provider Box, receives the TV signal. The raw EPG data is extracted from this broadcast TV signal (mobile or fix) and possibly from other internet sources. The PINGO box redistributes the EPG indoor for other mobile devices (phones, PMP) connected via WiFi.

In this scenario, there is no strong bandwidth constraint for the delivery of EPG data. Additionally, given the requirements that we want to minimize the mobile client developments and processing requirements, and we want to keep fine-grain synchronization between raw the EPG data and the presentation form, we decided to use a streaming delivery of EPG data in a presentation form. Given its mobile standard status, we selected the 3GPP DIMS format.

The EPG raw data is therefore transformed, based on templates, into a stream of presentation information which is updated regularly as new EPG data arrives. In this system, since the source of EPG data is mainly the broadcast channel, the EPG data is delivered in a push mode, using scene updates, as opposed to traditional AJAX pull mode.

The system functions as depicted in Figure 3. First, a designer creates two templates. The first template, called the Main Template, defines the general look and feel of the EPG. This template also handles the navigation between channels and display of information for the current channel. This template is initially empty of programs, but it identifies an entry point for the programs to be inserted, called the Event Dictionary. The second template, called the Event Template, defines the structure of a TV Event or TV Program. It is void of data but contains several entry points for the Event Information to be inserted (e.g. the name of the program, start time, duration, description ...). These two templates are written in the target presentation language. Once written, these templates are provided to a server.

The server (in the PINGO Box) is in charge of three tasks: first, sending the initial empty presentation; second, cloning the Event template, filling it in based on the EPG data it receives and sending the result in the form of a presentation update; and third, aggregating the presentation updates to provide a complete presentation for clients who have not joined the streaming session from the beginning or for those (e.g. Web browsers) who just support SVG (not DIMS).

The client in this system only implements a 3GPP DIMS player, it receives an initial scene which displays an empty EPG and then receives updates that progressively add new programs, replace the current time, or delete old programs. Figure 2 shows one possible scene structure. The scene is made of some Javascript code to create, at initialization, the layout of the EPG based on the terminal characteristics. It also contains the event dictionary on which updates are applied. When updates are received, some Javascript code is executed to determine if the program should be displayed based on the presentation time and the program time, on

previous user interactions and on the current EPG view. In this example, the whole EPG navigation or channel switching is realized using Javascript code.

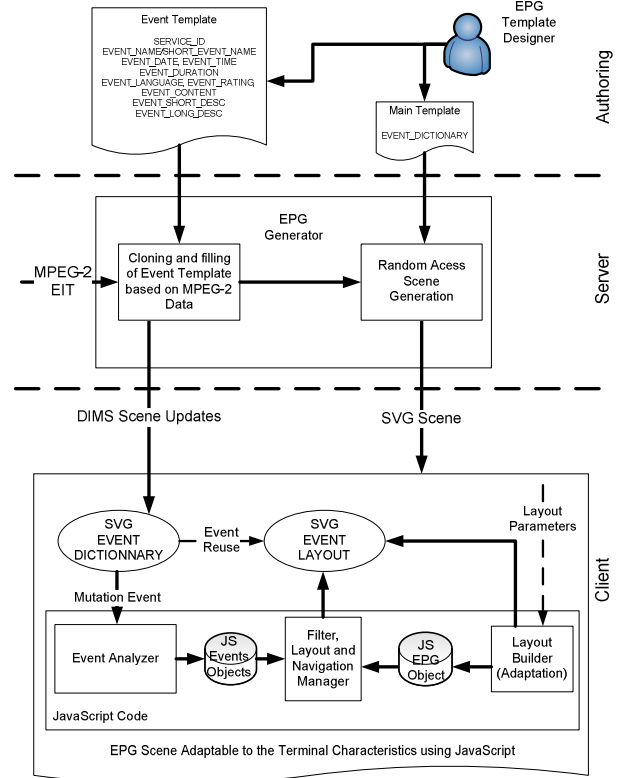


Figure 2 – Description of the end-to-end chain for the creation, delivery and presentation of EPG data in the PINGO System

4. DISCUSSIONS AND RESULTS

In this approach, a few points should be noticed. First, the description of a program is transmitted in the target presentation language (not in a dedicated metadata language), and then analyzed to produce the Javascript equivalent in order to be able to do client-side filtering and navigation. Programs are not described using some generic EPG XML format. This choice has been made to limit the client-side processing because otherwise if the program were sent using a generic XML format, the client would have had to create the SVG objects that represent the program using Javascript. Previous experiments showed us that the creation of objects with Javascript is slower than the creation of objects upon reception of scene updates. Similarly, in our system, we have tried to minimize the read/write access from the Javascript to the scene and we kept in Javascript only algorithmic operations (search, sort ...). Only minor attribute changes are done to the SVG scene using Javascript.

This system has been implemented and validated using the GPAC DIMS Player [13]. Figure 3 shows snapshots of the SVG/DIMS EPG produced by our system. Figure 4 demonstrates its usage of three mobile devices (SPV C500, Samsung i780 and Glofish V900) running Windows Mobile and the GPAC player. Similar experiments have been also made using the MPEG-4 BIFS language.

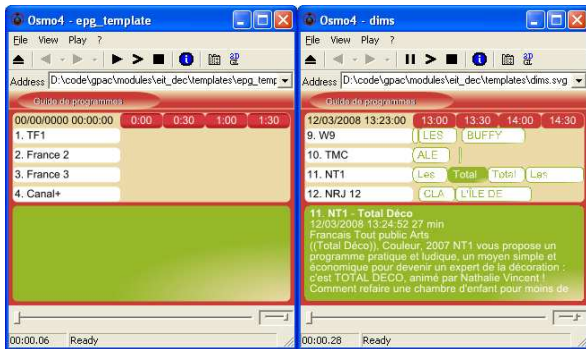


Figure 3 – Empty (left) and filled (right) PINGO Electronic Program Guide

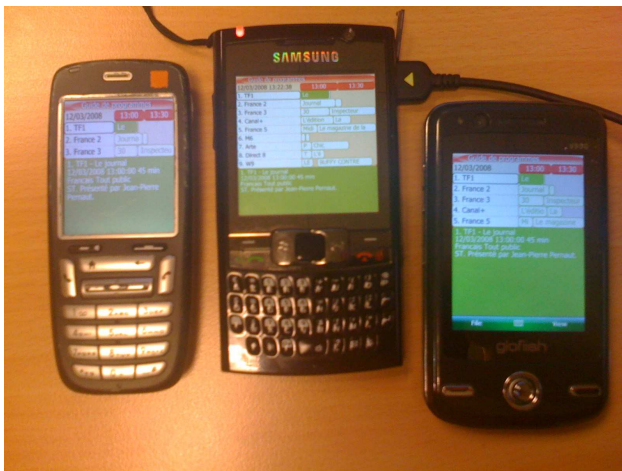


Figure 4 – Adapted display of the PINGO Electronic Program Guide on three different mobile devices

5. CONCLUSION

In this paper, we have exposed the existing approaches for the generation and display of electronic program guides for interactive TV. We have discussed the pros and cons of these approaches. We have presented our scenario and argued for the need for an approach based on streamable presentation languages. This approach allows separating the decoding or interpretation of raw EPG data from their presentation, thus allowing the display of EPG on constrained devices. In our system, the use of declarative presentation languages enables the designer control over the presentation and its adaptation to the device characteristics. We believe that this approach, applied here to EPG, is actually very generic and could be applied to the presentation of any type of metadata, through the use of templates. In future work, we will investigate several improvements to this system: the ability to have more input formats (web feeds), to support more output formats (XHTML). We will also work on improvement of the adaptation features of our prototype.

6. ACKNOWLEDGMENTS

This work has been partly financed by the French General Directorate for enterprises and the French Business Cluster for Digital Content.

7. REFERENCES

- [1] "Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems", ETSI EN 300 468 V1.7.1
- [2] "Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Electronic Service Guide", ETSI TS 102 471 V1.2.1
- [3] ANT Fresco Browser, <http://www.antlimited.com>
- [4] "Dynamic and Interactive Multimedia Scenes (DIMS)", 3GPP TS 26.142, <http://www.3gpp.org/ftp/Specs/html-info/26142.htm>
- [5] "Information technology - Coding of audio-visual objects - Part 11: Scene Description and Application Engine", ISO/IEC IS 14496-11
- [6] "The XMLHttpRequest Object", W3C Working Draft 15 April 2008, <http://www.w3.org/TR/XMLHttpRequest/>
- [7] Joost Widget API, Runtime Engine Information, <http://dev.joost.com/widgets/api/engine.html>
- [8] The XML TV Project, <http://wiki.xmltv.org/index.php/XMLTVProject>
- [9] "DVB Data Broadcasting Implementation Guidelines", ETSI TR 101 202 V1.2.1
- [10] "File Delivery over Unidirectional Transport", RFC 3926, <http://www.faqs.org/rfcs/rfc3926.txt>
- [11] "Broadcast and On-line Services: Search, select, and rightful use of content on personal storage systems (TV-anytime)", ETSI TS 102 822
- [12] "Information technology - Multimedia content description interface - Part 1: Systems", ISO/IEC IS 15938-1
- [13] Le Feuvre, J., Concolato, C., and Moissinac, J. 2007. GPAC: open source multimedia framework. In Proceedings of the 15th international Conference on Multimedia (Augsburg, Germany, September 25 - 29, 2007). MULTIMEDIA '07. ACM, New York, NY, 1009-1012.
- [14] Cruickshank, L., Tseklevs, E., Whitham, R. and Hill, A. 2007. Making Interactive TV easier to use: Interface design for a second screen approach. The Design Journal, 10, 3.
- [15] Lee, H., Yang, S.-J., Kim, J.-G., and Hong, J. 2005. Personalized TV services based on TV-anytime for PDR. In Proceedings of the International Conference on Consumer Electronics (January 8-12, 2005). ICCE 05, 115-116.
- [16] Harrison, C., Amento, B., and Stead, L. 2008. iEPG: an ego-centric electronic program guide and recommendation interface. In Proceeding of the 1st international Conference on Designing interactive User Experiences For TV and Video (Silicon Valley, California, USA, October 22 - 24, 2008). UXTV '08, vol. 291. ACM, New York, NY, 23-26.
- [17] Cesar, P., Bulterman, D. C., and Jansen, A. J. 2008. Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content. In Proc. of the 6th European Conference on Changing Television Environments (Salzburg, Austria, July 03 - 04, 2008). M. Tscheligi, M. Obrist, and A. Lugmayr, Eds. Lecture Notes In Computer Science, vol. 5066. Springer-Verlag, Berlin, Heidelberg, 168-177.

Automatic Recommendation of (IP)TV Program Schedules using Sequential Pattern Mining

Shinjee Pyo

Depart. of Info. & Comm.
KAIST

119, Munji St. Yuseong-gu, Daejeon
+82-42-866-6246

sjpyo@kaist.ac.kr

EunHui Kim

Depart. of Info. & Comm.
KAIST

119, Munji St. Yuseong-gu, Daejeon
+82-42-866-6249

lins77@kaist.ac.kr

Munchurl Kim

Depart. of Electrical Engineering
KAIST

119, Munji St. Yuseong-gu, Daejeon
+82-42-866-6137

mkim@ee.kaist.ac.kr

ABSTRACT

In the conventional TV viewing environment, the numbers of TV channels and program contents are mostly limited so that the accessibility to TV program contents is made by manually hopping the TV channels and selecting the preferred TV programs. However, with the advent of (IP)TV services with the excessive number of TV program contents via many TV channels, TV viewers (users) are required to make much effort to choose their preferred TV channels or program contents, which becomes much cumbersome to the users. Interestingly, user's behavior on TV watching exhibits sequential patterns on the consumption of TV program contents in somewhat regular manners. Therefore, in this paper, we present an personalized automatic TV program scheduler by making reasoning on users' TV viewing patterns from the usage history of the watched TV program contents using sequential pattern mining so that accessibility of users to TV program contents can be made by requiring much less effort by users.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Application – *data mining*.

General Terms

Algorithms

Keywords

Sequential pattern mining, Recommendation, IPTV, Program schedule, Data Mining

1. INTRODUCTION

Recently, a new media service such as IPTV has started so that tremendous amounts of TV program contents become available at the user sides. Also, users can access many TV program contents in each IPTV service or via many TV channels. Therefore, research on automatic TV program recommendation has been made to provide easy access to the TV programs for users to alleviate the burden in finding their preferred TV program contents. On the other hand, the TV watching behaviors of users exhibit somewhat regular TV watching patterns specific to particular users in chronological and/or semantic perspectives. This means that the watched TV program contents are semantically or chronologically related. This can be easily found in goods purchases in which the purchased items

are related sequentially in a period of time. Based on this characteristic of users' goods purchases, many researchers have studied to analyze users' purchase history and extract the sequential purchase patterns using sequential pattern mining technique [2]. Similarly, TV viewers may also have their regular patterns in TV watching in terms of their watched TV program contents from chronological perspective on each date. In this paper, a scheme on automatic and personalized TV program schedule recommendation is studied so as to make it easier to watch (IP)TV program contents sequentially. By providing a sequence of TV program contents to a user as a personalized TV program schedule, the user can capture in advance the information of upcoming TV programs which are interesting to him. It is believed based upon the authors' best knowledge that the concept of automatic and personalized TV program scheduler recommendation be first studied in this paper, although many recommendation schemes of TV program contents have been studied in the literature. The proposed automatic TV program scheduler recommendation is based on sequential pattern mining in conjunction with collaborative filtering. The rest of this paper is organized as follows: Section 2 discusses our proposed IPTV schedule recommendation scheme, Section 3 shows the experimental results and Section 4 concludes the paper and discusses about future works.

2. Proposed System for IPTV Schedule Recommendation

2.1 System framework

Figure 1 shows a recommendation system of our proposed automatic (IP)TV program scheduler that consists of an (IP)TV server, a recommendation engine and user terminals. The (IP)TV server controls users' activities such as log-in, log-out, registration and pushing play/stop buttons, and records the users' usage history such as the titles, and start-/end-times of the watched TV program contents for each user. The usage history of the watched TV program contents will be used as important clues for personalized recommendation for each user. The recommendation engine (1) clusters the users as similar user groups using each user's usage history, (2) sequentially extracts the watched TV program contents in a chronological order using sequential pattern mining for each similar user group, (3) extracts the user's preferences on TV channels and program contents and (4) generates the personalized TV program schedule using the preference and extracted TV program patterns. When a user logs in an (IP)TV user terminal, then the IPTV server recognizes the

log-in user ID and transfer the user information and the log-in time to the recommendation engine to request the recommendation for the user. Then the engine finds the user's appropriate recommendation schedule according to the log-in time and the user ID, and then IPTV server sends the TV program schedule recommendation to user's terminal. The usage history is stored in user's usage history database as user's profile information. If the user feels the recommended TV schedule is acceptable, then the TV program contents are consumed in the user terminal in a recommended sequence order. The IPTV server requests to the content server the TV program contents as scheduled.

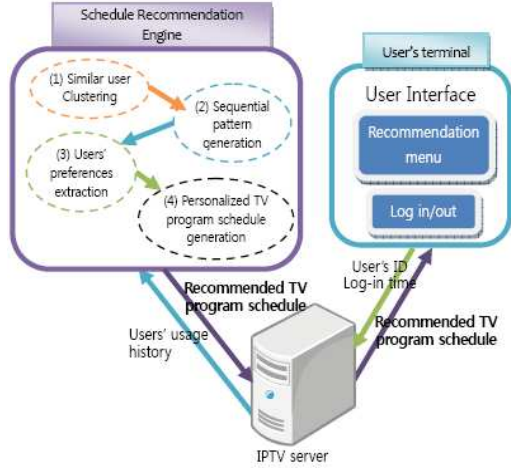


Figure 1 A Recommendation system for personalized IPTV program scheduler

2.2 Recommendation Procedure

2.2.1 Grouping Similar Users

For training, the data of TV program viewing history is used which had been collected by ACNielsen Korea Research Center and was recorded for 2,522 people from December, 2002 to February, 2003. The first step for personalized IPTV schedule recommendation is to cluster TV viewers into similar user groups. In this step, we clusters the TV viewers into similar groups for which more semantic collaborative filtering can be possible against an active user. Here, the active user means the user who logged in the TV terminal. For recommendation of TV a program scheduler to the active user, a sequence of TV program contents are constructed with more frequently watched TV program contents in the most similar user group to the active user. For similar user grouping, a feature is defined in terms of the N most interesting TV program contents for the active user. The interest on a TV program content is measured by TF-IDF (Term Frequency – Inverse Document Frequency), which reflects the watching duration and emission times of the TV program content. The vector space model is usually used to measure the similarity between text documents as vectors. It is used in information filtering, information retrieval and relevancy rankings. In the system, we use the vector space model to compute the similarity between a user and the active user based on the feature vector which represents user's TV watching preferences. The similar measure in the vector space model is defined as follow:

$$\cos \theta = \frac{\mathbf{v}_{u_a} \cdot \mathbf{v}_{u_i}}{\|\mathbf{v}_{u_a}\| \|\mathbf{v}_{u_i}\|}, \mathbf{v}_{u_a} = [w_{1,u_a}, w_{2,u_a}, \dots, w_{N,u_a}] \quad (1)$$

where \mathbf{v}_{u_a} and \mathbf{v}_{u_i} are the feature vectors of the active user (u_a) and a user (u_i) in the usage history data. w_{i,u_a} represents the i -th ranked interesting TV program content by the active user. For the experiments, we use the usage history data of the watched TV program contents for 50~55 year-old women who have much less data sparseness. By comparison of users to the active user according to Eq. (1), we select the similar taste user group of the active user with the top M similar users according to Eq. (1). In the experiments, M and N is set to 30 and 10.

2.2.2 Sequential Pattern Mining

In this section, we extract sequential patterns (a sequence of TV program contents in order of time) from the usage history data of the watched TV program contents of the similar user group. First, we divide the usage history data into different days of the week and select frequently watched TV program contents for each day of the week. The frequently watched TV program contents are selected based on their accumulated watching lengths which exceed a predefined threshold. Therefore, the more frequently watched is a program content in the similar taste user group, the larger the accumulated watched length of the program content is. This is different from the traditional sequential pattern mining in which the frequency of item purchase is only considered when selecting the frequent item. However, in the TV watching environment, the watched time length for each program content is also important. So, we consider the quantitative information of user's TV program watching time. The sequence consists of the program contents within a date. If we index the program titles as numbers, watching time length (q_{j,u_i}) of a TV program content (p_j) by a user (u_i) can be represented as follows:

$$q_{j,u_i} = \frac{wt_{j,u_i}}{d_j} \quad (2)$$

where d_j is the broadcasting duration of program p_j of a user u_i and wt_{j,u_i} is the watching duration that the user u_i has watched program p_j . The watched TV program contents for which q_{j,u_i} exceed a threshold value constitute a timed sequence as shown in Table 1.

Table 1. Quantitative timed sequence

user ID /day/date	Quantitative timed sequence
u1/Sun/2002.12.1	$\langle P_{134,0.5}, P_{56,0.2} \dots \rangle$
u1/Sun/2002.12.8	$\langle P_{578,1}, P_{134,0.6} \dots \rangle$
u2/Sun/2002.12.1	$\langle P_{134,0.4}, P_{87,0.4} \rangle$

For the timed sequences for all users in the similar taste user group, we accumulate all q values for the same program as follow:

$$Q_{pi} = \frac{\sum_{k=1}^N q_{i,u_k}}{B_{p_i}} \quad (3)$$

q_{i,u_k} is a watched time length of program p_i of user u_k and B_{p_i} is a broadcasting frequency of p_i . N is a total number of users in the user group. The Q_{p_i} contains the quantitative information of all users in similar user group. This value represents a royalty on a specific program p_i of all users. Q_{p_i} is maximum when all users in the user group fully watched the program p_i every emission time. So, the maximum value of Q_{p_i} is $N(=30)$ and the minimum value of is 0 which is the case that none of the users watches it. We set the threshold value for Q_{p_i} to be 2 as the minimum support value to construct frequent items. In the history database divided by day of week, we can select the frequent programs of day of week using the Q_{p_i} of each program and minimum support value. In order to make a sequential pattern based on the selected frequent program contents, we use the PrefixSpan algorithm [2]. First, we find sequences having the frequently watched program as prefix, so only the sequences which contain the frequently watched program should be collected. Second, in the sequence containing the frequently watched program, only the subsequence prefixed with the first occurrence of the frequently watched program should be considered. This process is called prefix projection. For example, in the program sequence $\langle p_{167} - p_{30} - p_{254} - p_{16} \rangle$ which contains the frequently watched program p_{30} , only the subsequence $\langle p_{254} - p_{16} \rangle$ should be considered for extracting sequential patterns having p_{30} . So, the projected sequence only contains the later part and the projected database of the frequent program consists of the projected sequence of the frequent program contents. In the projected database of the frequent program contents, we select the frequent program contents using the minimum support value (threshold value) of Q_{p_i} . Through the process for selecting the frequent programs in the projected database of the frequent program content for each date of the week, we can find the length-2 sequential pattern. A length-2 sequential pattern is expressed as " $p_i - p_j$ ". p_i is a frequently program content on a date and p_j is a frequently watched program content of the projected database of p_i . The following table shows the length-2 sequential patterns.

Table 2. Total 593 length-2 sequential patterns

day	length-2 Sequential patterns
Sun	$\langle p_{528} - p_{617} \rangle, \langle p_{528} - p_{366} \rangle \dots 263$ patterns
...	
Sat	$\langle p_{54} - p_{593} \rangle, \langle p_{528} - p_{54} \rangle \dots 160$ patterns
Total 593 length-2 sequential patterns	

For the projected sequence database, a new projection can be made based on the newly found frequently watched program contents in a recursive manner. Therefore, by repeating the projection and selection of frequent programs in the projected database, we can make longer sequential pattern. In this paper, we generated up to length-3 sequential patterns. The following table shows the extracted length-3 sequential patterns.

Table 3. Extracted length-3 sequential patterns

day	length-3 Sequential patterns
Sun	$\langle p_{528} - p_{617} - p_{590} \rangle, \langle p_{528} - p_{366} - p_{180} \rangle \dots 1743$

...	
Sat	$\langle p_{528} - p_0 - p_{54} \rangle, \langle p_{528} - p_0 - p_{598} \rangle \dots 909$ patterns
Total 3181 length-3 sequential patterns	

2.2.3 Extracting User Preference

In this section, we extract program, genre and channel preference of each user. Program preference for each user is computed by Eq. (2). Using Eq. (2), the top 20 programs are selected for a user's favorite programs. Other preferences are considered for the personalized program recommendation, such as genre preference and channel preference. We extracted genre and channel preferences based on the normalized watched time length of each genre and each channel by the total eight genres and six channels.

2.2.4 Personalized TV Schedule Recommendation

There might exist many sequences of having a same frequently watched program content followed by different frequently watched program contents. In this case, we need to choose one sequence that is more preferred to the active user. For this, a rank model is defined to select one of the candidate sequences and is given by

$$RS_{u_k,i} = (1 + PV_{u_k,i}) * (1 + GV_{u_k,i}) * (1 + CV_{u_k,i}) * Q_{p_i} \quad (4)$$

$RS_{u_k,i}$ is the rank score of program p_i . $PV_{u_k,i}$, $GV_{u_k,i}$ and $CV_{u_k,i}$ are the preference values of program, genre and channel, respectively. Using Eq. (4), we recommend the sequential patterns with largest RS values.

3. Experimental Results

For the experiments, we use two sets of training and test data, each of which consists of the watching history data from March, 2003 to May, 2003 for the users in the same group. In the experiment, the proposed scheme is tested with the usage hostage of the watched TV program contents which has not been seen during the training phase. First, we compare group to group by checking the length-1, length-2 and length-3 extracted sequential patterns in the test data. Second, we validate our personalized recommendation of the TV program scheduler based on sequential patterns using each user's preference by checking the each user's test data. Third, we compare the recommendation precision of our proposed sequential pattern mining with the existing PrefixSpan algorithm.

3.1 Recommendation Performance

In the training data of a similar user group, we generated length-1, length-2, and length-3 sequential patterns (frequent programs) using the proposed sequential pattern mining. So, for test, we check whether these recommended sequential patterns are really watched by comparing with the test data. Figure 2 shows the recommendation precisions for the length-1, length-2 and length-3 sequential patterns. As shown in Figure 2, the recommendation precision of length-1 sequential patterns of each day of the week is close to 1. However, as the length of the recommended sequential patterns increases, the recommendation precisions become lower. Also, recommendation precisions of each date of the week are different because the watching behaviors of users are different each date.

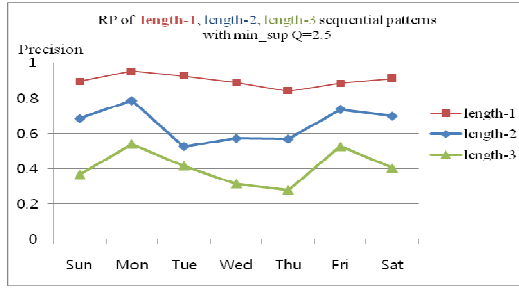


Figure 2 Recommendation precisions for recommended sequential TV program of length-1, length-2 and length-3

3.2 Recommendation Performance with user preferences

For the projected program sequences based on frequently watched TV contents, we obtain length-2 or longer sequences starting from FWP items. Instead of recommending one or more projected sequences for a same FWP item, only one projected sequence is selected based on user preference on program items in this experiment. So, when we recommend the first sequential program pattern among the generated sequential patterns using the rank score value, the precision is measure by evaluating whether the user has also watched the recommended sequential program patterns in the test data of usage history. Figure 3 shows the recommendation precision of length-2 and length-3 sequential patterns based on user preference. As shown in Figure 3, the performance on recommendation precision of length-2 sequential patterns exhibits approximately 60% ~ 40% over the weeks and for recommended sequential patterns of length-3, about 22%~50% of precisions are obtained.

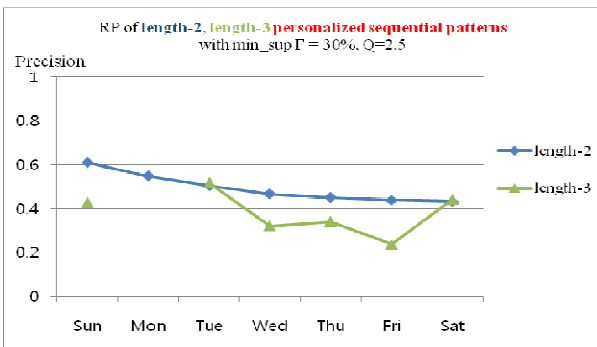


Figure 3 Recommendation precisions of personalized sequential patterns

The performance of recommendation of sequential patterns based on user preference is lower is because only one projected sequential pattern is recommended for each frequently watched program item. Also, there is a data sparseness problem. As shown in Figure 3, the usage history for the length-3 sequences does not exist on Monday.

3.3 Comparison of performance between our proposed sequential pattern mining and original PrefixSpan

We compare the proposed method with the original PrefixSpan algorithm in terms of recommendation precision for length-1,

length-2 and length-3 sequential pattern (frequent program). The minimum support of watching frequency for both algorithms is 20% of total number of sequence, and the minimum support of for proposed algorithm is 2 for selecting frequent programs of each day and 20% of the Q_{p_i} . As shown in Figure 4, square dotted and circle dotted line represents precision of proposed algorithm and PrefixSpan algorithm, respectively. All precisions of the proposed algorithm for length-1, length-2 and length-3 sequences are higher than those of PrefixSpan algorithm. This experiment shows the superiority of the proposed algorithm over PrefixSpan algorithm.

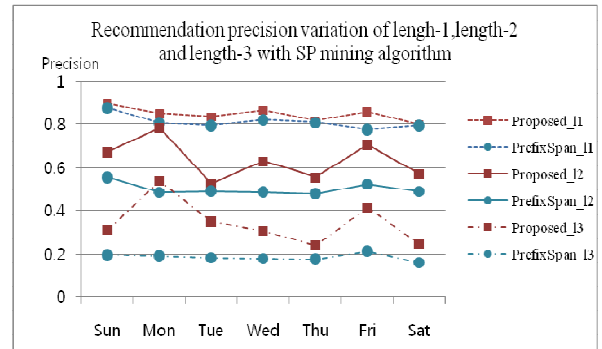


Figure 4 Comparison of recommendation precisions of length-1, length-2 and length-3 sequential patterns

4. Conclusion and Future works

In this paper, we propose an automatic and personalized TV program schedule recommendation scheme based on sequential pattern mining. To recommend personalized TV program schedules to a user, first we group the similar user groups, extract the sequential patterns and user's preferences and associate them to generate the personalized sequential patterns for more meaningful as recommendation candidates. Using the generated sequential patterns with the features of accumulated watched time lengths, and user preferences in addition, we achieve improved recommendation performance against the PrefixSpan method. As our future works, we plan to extend the scheduler recommendation for the content service for a non-real time case of VoD applications. Also, we will consider the users' feedback into our TV program scheduler recommendation to design an elaborate and adaptive recommendation system.

5. ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MKE/IITA. [A1100-0801-3015, Development of Open-IPTV Technologies for Wired and Wireless Networks].

6. REFERENCES

- [1] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," In Proc. 2001 Int. Conf. Data Engineering, Vol. April 2001.
- [2] Baoyao Zhou, Siu Cheung Hui, Kuiyu Chang, "An Intelligent Recommender System using Sequential Web Access Patterns," In Proc. 2004 IEEE Conference on Cybernetics and Intelligent Systems, December 2004.

New potentials for T-Learning and informal learning with ‘mashuped Television’ – reinventing of Hypermedia?

Robert Strzebkowski
Beuth University of Technology Berlin
Luxemburger Str. 10
13353 Berlin (Germany)
++49-30-4504 5212
robertst@beuth-hochschule.de

Björn Münchau
meta-morph.tv
Fidicinstr. 3
10965 Berlin (Germany)
++49-30-202 35 925
bjoern.muenchau@meta-morph.tv

ABSTRACT

In this paper we discuss some potentials for informal learning in new TV based interactive environments based on IPTV and WebTV. The current technologically, functionally und usage pattern developments regarding the increasing convergence between the television and the Internet rise new possibilities for TV media based Learning. We reflect these new changes on current media usage patterns, on some crucial web 2.0 aspects as well on important findings in the area of e-learning. Two examples from the own work in this area will be short discussed.

Categories and Subject Descriptors

D.H.1.2 User/Machine Systems, H.3.3 Information Search and Retrieval, H.5.1 Multimedia Information Systems, H.5.2 User Interfaces, H.5.4 Hypertext/Hypermedia

General Terms

Management, Documentation, Design, Economics, Experimentation, Human Factors, Standardization, Languages, Theory.

Keywords

iTV, IPTV, WebTV, Web 2.0, TV 2.0, T-Learning, TV-based Learning, Video centered Learning, hypermedia, mashup, informal learning, connectivism, e-Learning 2.0

1. INTRODUCTION

At present we experience a dramatically change in the distribution, consumption, production, exchange and usage patterns of digital media and information: “the explosion in broadband uptake combined with the relentless space of technological innovation is driving a major change in consumer behaviour and is transforming our traditional media landscape” [22]p.3. During the last three years the broadband connections in Europe grew about 95% from 44 million households to over 85 millions and 120 millions subscribers! [14], [40]. The enormous increase of the Internet infrastructure, the large accessibility to the broadband Internet – wired as well wireless – for all relevant user groups as well the exponential growth of the digital media and telecommunication devices – forecast of 19 Billions! worldwide by End of 2010 [6] – lead to the evolution of *connected society*. The ubiquitous connectivity and access to media content and information as well the seamless possibility to communicate, to create and to publish new content drive the use of web 2.0 applications and the consumption of digital media.

For instance the use of so-called ‘social media’ increased worldwide between 2006 and 2007 over 66%. 50% of the worldwide Internet user use regularly web 2.0 applications [27]. Ca. 45% of the Internet user are user of blogs and between 56% and 60% of the Internet user visit social networks [36]. 72% of the worldwide Internet user watch regularly videos online [27] and each day nowadays will be uploaded near 150,000 videos to YouTube video portal [41]. In the US are being viewed monthly about 10 Billions videos online! [32].

Beyond of this impressive usage data there are some aspects, which are crucial for the current changes in the media, information and telecommunication landscape.

One of the major aspects is the strong growing *convergence* regarding digital media eco systems, digital media technologies, applications, digital media formats and usage patterns of the new convergent systems [5]. The aspect of the convergence is very often based on the second factor, on the *connectivity* and the *mashup* approach. Started by the seamless connection of different media devices through connected people until the mashups of online services, applications and media content items. *Connected Entertainment* is here a new current ‘buzzword’ [22]. These last aspects have strong relation to the following factors: *device independence*, *mobility* and ‘*information atomization*’. The latter means the dynamic compilation of content offer is based on content ‘chunks’ from diverse sources – documents, media objects as well servers and providers. Last but not least are the ‘*social components*’ very important in the new media landscape. Keywords like *community*, *collaboration*, *content production as UGC*, *immediately reflection* as meaning exchange or feedback should be here representative.

The daily use of IP-based media and web 2.0 applications with the mentioned aspects above influence our patterns of media consumption, information collection and exchange, learning and working in the private as well in the vocational environment.

The extremely growth and amount of information as well their increasing short endurance regard a permanent life long learning process. *Informal* or so called *just in time learning* become here more importance in private as in vocational scope.

These short noted key aspects above about the current developments regarding the technology and the use of digital/IP media as well the requirements in the area of learning deliver an important conclusion base for the following considerations about TV media based learning – T-Learning.

2. TV, Internet and the information genre

TV and Video - still the most important media and media formats

Despite of the exponentially growth of the use of Internet TV like or digital video based media are at present still the most used electronic media or media presentation form by the majority of media users [16]. Included in this view are the traditional TV, interactive TV (iTV), IPTV and WebTV/Internet video platforms. Nowadays there is being made a distinction between IPTV as based on closed networks and presented on TV sets and WebTV as running on open internet and delivered over a PC interface, but already now is starting the increasingly fusion between the both platforms [31], [32]. The average consumption time of the traditional TV broadcasts is either remaining or even increasing [15]. A closer look at the specific age levels of the viewers reveals the trend among young people between the ages 14 and 29, who are spending increasingly more time on the Internet than on using TV [26]. However, at the same time those people are building the main user and viewer group of the currently exponentially growing WebTV offers in the Internet [32]. 92% of the young Internet users in Germany – between 14 and 19 years old – are using regularly live or time-shifted video or TV broadcasts online. The demand for the use of video by all German Internet users increases in the time between 2007 and 2008 from 45% to 55% [1]. In Europe “since September 2006 the number of online users who have watched a video over the Internet has increased from 31% to 82% in Jan 08 – an [over 150%] increase” [22]. The growth of the IPTV sector is running distinct slower than the one of WebTV but the development is also very respectable. In Germany there was a user increase of ca. 200% between 2006 and 2008! [31].

Conclusion: because of the successful development of the WebTV/IPTV infrastructure as well of the enormous demand, interest and popularity of TV and video – still as traditional TV or as online media via IPTV or WebTV – this media seems to be an important communication and motivation ‘channel’ for information and learning. Especially for the younger users as heavy Internet as well WebTV users open the TV like media additionally information and learning opportunity.

Information - a crucial genre by Internet as well as by TV

There are five general experience factors - usage needs and gratifications - for the use of Internet and Television: *orientation* (information and learning content), *relaxation* (fun, suspense), *compensation* (from the day problems distracting content), *amusement* and *social experience* [26]. Regarding the factor orientation is the Internet especially by the young users – 14 – 29 years old - in advantage. The 46% of the German media users in the wider age range between 20 and 59 mean the TV is still better suited for the information aim than Internet and only 19% vote here for the Internet [37]. For 62% of all German Internet users is the primer usage need of Internet the access to the information [1]. Ca. 40% of the consumed online videos (WebTV) belong to the genre news/information and stay so on the first place of the used categories. For the significant majority of all media users is still the TV – traditional or IP-based - the reliable information source.

Recent statistics regarding TV offering as well TV consumption show, that the category/genre information is still noticeably increasing [42]. 60% of the offered TV program by the public broadcasters and 40% by the private in Germany is information. By the latter grew this category between 2005 and 2007 between 50% and 100%! 36% of the daily viewing time of the German TV viewer concerns the category information [42]. The information

category comprises news, documentaries, magazines, talks, sport, advisory, teaching broadcasts or scientific/technician presentations as well other similar formats.

In the ‘traditional’ digital TV landscape one can observe the increase of special interest channels and lot of them deal with the information genre. For instance on the BSkyB digital TV platform in England there are numerous information channels like Discovery Science, Discovery Wings, Discovery Civilisation, Biography, Nat Geo or BBC News 24, SKY News, etc. The BBCi interactive TV service counts weekly 11 Mio viewers [BBCi [4]. An onother interesting trend are WebTV channels based on certain broadcast programmes. In Germany is here the interesting example “Welt der Wunder TV” (World of the Wonder TV – www.wdwiptv.de), as a program on RTLII and as an own WebTV channel.

In the scope of WebTV - beside the big video sharing and video clip portals like YouTube, Metacafe or the like - there is worldwide a respectably growing range of information related WebTV channels and programs (e.g. de.wwitv.com). Especially the traditional TV broadcasters as well the newspaper publishing companies are strengthened using WebTV platforms for distributing their TV or publishing content online and are thus strong contributing to the effect of widening of TV like information programs and channels. Good examples here are the BBCi Websites and the successful BBC iPlayer with monthly 15 Mio visits [4]. Also in Germany the both big public TV broadcasters ARD and ZDF offer an enormous amount of TV information content – video as well as interactive multimedia applications - in their ‘Mediatheks’. The ZDF Mediathek counts ca. 10 Mio visits monthly! The newspapers like The Guardian, Der Spiegel or Stern have a comprehensive offer of video content but mostly purchased or syndicated from other video or TV media producers.

Conclusion: the evident increase of the offer as well of the interest in the scope of the genre information or news in TV like media build substantially content base for learning purposes. Moreover this reflection was necessary for the evidence on strong represented information genre also as WebTV content. Very often there is a naive argumentation especially regarding WebTV as only an ‘Entertainment Tube’.

3. Learning with TV like media – T-Learning

Some aspects of learning processes with digital media

TV as well video provide very effective presentation form for learning content and for didactic-pedagogical purposes. Seels [35] examines many research programs and findings from over the last 40ty years in the area of learning effects with Television and concludes a quite positive picture. She confirms the complexity concerning the ‘measuring’ effects of TV based learning and discusses also some negative effects. Especially the research in the 1970ies and in the 1980ies has delivered some interesting and important findings about AV media based learning. For instance the Supplantation Concept from Salomon [33] or the Aptitude-Treatment-Interaction (ATI) Approach [7]. This former research was based on the *Cognitivism* theory as well on the information-processing approach and has focused on different media attributes related to the different and individual learner characteristics. In particular the research in the area of Instructional Message Design [13], [24], [23] has identified many effective presentation forms and their combinations as well interdependences for positive learning effects.

A successful learning with TV like content - despite the audiovisual and dramaturgic quality - depends on many different factors like the learning purpose, need and motivation, the social and didactical integration in and of the learning process, the conscious reflection about the viewed content as well the degree of active self selection of the learning content [35]. In particular is the aspect of active engagement with the learning object and moreover the active and self-regulated use of media and the media content to create individualized knowledge base concerning the individual requirements and motivations the crucial factor for a learning impact [21]. Those aspects are well explored in the debates on constructivism [11] as well on hypermedia and multimedia learning environments [29]. The main attribute by the media design is here the form and the degree of interactivity [34], [20].

The interactive functionalities and services based on the permanently broadband online connection shape the main difference between the current and the older educational approach of TV like media. First with the serious advent of interactivity by the Television as well with the increase offer of TV like media (WebTV) in the Internet are returned the ideas and demands for T-Learning.

Characteristics of T-Learning

There are already very appropriate definitions of T-Learning in some earlier papers like by [2], [30], [39]. One can realize the gradual changes in these definitions during the last years concerning the coalescence of TV and Internet/IP-based media technologies as well the increase importance of the convergence. Atwere & Bates [2] underline in their definition from 2003 the TV set as a primer delivery device for T-Learning: "T-learning (television-learning) learning with interactive TV or similar screen-based device with video-rich content delivered via one or a number of different platforms (satellite, cable, digital terrestrial) but not primarily learning with a personal computer as it is recognised today". Aarreniemi-Jokipeltö (2005) [28] extends this characterization: "T-learning is seen as a convergence of cross media and e-learning [...] television or device suitable for viewing broadcasted contents is the primary medium in T-learning [...]" Interactivity, Video-On-Demand, enhanced TV services, stretched content, convergence and e-learning are the crucial aspects in the former definitions. Moreover Atwere & Bates [2] have quite early recognized the importance of personalization approach, which plays nowadays under the keyword targeting increasingly importance.

Almost all of the existing definitions include the weighty aspect of e-learning and set T-Learning as a subset of it [2], [28], [39]. This classification view equates to the common definitions of e-learning [18]. As mentioned in the introduction of this paper the current changes in the digital media and communication technologies as well their usage patterns influence significantly our private and professional live as well the form of acquiring information and learning. Hence e-learning underlies also to these changes, which will be sometimes sub summarized as e-learning 2.0 [3], [12]. Three aspects from the scope of e-learning 2.0 are important for the subject of this paper: the increasing significance of informal learning, the connectivity and mashup approach between different content items as well the social and creative Web2.0 interactivities.

Informal learning occurs independent from an organized learning process with prior defined learning objectives and outcomes. This form of learning is dominating the human life up to 80%! The characteristic aspects are here immediacy and relevancy as well meaningful and self-directed learning process based on individual

interest on a certain subject [8]. In the area of e-learning plays informal learning increasingly considerable form of learning strategy because it is directly connected to the current demand for knowledge mostly applied in a problem or task solving situations in the vocational practice [9]. In the scope of this paper is T-Learning considered as a learning process in informal – private – situations and not in a vocational environment. This is the one similarity between informal and T-Learning. The second one is the aspect of immediacy and self-motivation as well self-directedness. The consumption of TV and TV like media is guided through an individual motivation and interest for a certain subject and not through a strong learning objective. A recent Motorola Survey (2008) [25] revealed that young person – 16 – 27 years old – are very interested "[...] to be able to interact (vote, get information, or purchase) with [the] favourite TV shows while watching them" – 57% – and „[...] to interactively learn about (and possibly purchase) items featured in TV shows“ – 68%. These statements confirm the importance and the demand for self-directedness regarding the content selection as well the content complexity [22]. Such interactive techniques can foster the degree of attention and a more active elaboration during the informal learning process.

Connectivity – the advent of 'mashup' Television

The demand for immediate access to the additional context sensitive information related to the showed TV content - whether about a learning subject or a product - requires the implementation of hypermedia content connections. The usual way to provide a hypermedia content system is based on the connection to the online distributed databases with metadata as well to the content objects. The dynamic composition of the linked and requested content and applications from different national or international distributed server to a one homogeneity looking interactive information or learning application establish a quite new form of online connected T-Learning systems. This development provide a quite good analogy to the learning approach of *Connectivism*, which "focuses on the inclusion of technology as part of our distribution of cognition and knowledge. Our knowledge resides in the connections we form - where to other people or to information sources such as databases" [38]. This approach meets the currently problematic of the exponentially information growth, their short actuality as well of the possible information overload. The solution here is to let the information externalized and first through an on-demand search and request will be collected the information 'chunks' together to a meaningful information. "We human exist in networks", "Learning consists of making and maintaining better connections to our networks" [8] p.7.

Self-directed learning through exchange and creative contribution

An active participation in a discussion during a learning process is one of the most powerful didactic methods for individualized learning. The articulation of arguments, concepts and ideas is the best way to active elaborate the acquired knowledge and to self proof the understanding of a subject as well in relation to the statements of other learning persons if there are lacks in the view of the learning subject. The constructivists has extensively investigated this aspect [11]. Nowadays there are great opportunities with the 'social tools' from the 'Web 2.0 Toolbox', which let the learner active exchange with each other. And as we have reflected in the introduction those 'tools' are being very willingly used from about 50% of all Internet users! Downes 2007 [10] has postulated the notion of 'e-learning 2.0' for the integration of Web2.0 techniques in e-learning processes. Big companies for instance are increasingly implementing e-learning 2.0 environments [17].

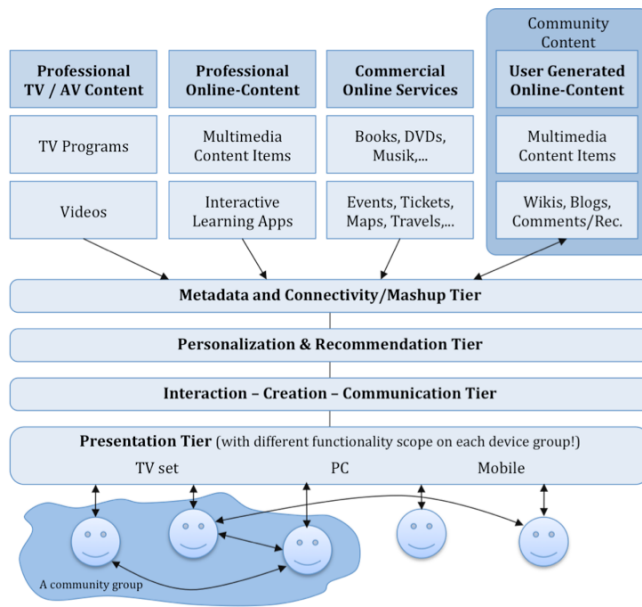


Figure 1: Model of a T-Learning environment with the focus on mashup/hyperlinked content, device independence and active content contribution of the subject community – TV2.0

The active role as a participant in an e-learning community is not constraint to post only text messages. In particular for the scope of T-Learning are the possibilities to submit self-created media like photographs, pictures, animations or videos very interesting. The survey from Microsoft [22] reveals the strong demand of the ‘Generation Digital’ for online media “where they can be in control and have an instant say, or share their views on the information they are viewing”. These aspects of active contribution can foster the attention and the motivation to engage deeper with a subject matter as well the building and growing of active special interest groups and communities [3], [11], [12], [17], [20].

Currently we work on two IPTV/TV2.0 T-Learning projects/scenarios – ‘Fernweh’ and ‘Home-Vision.TV’. The first one is a travel TV channel with context sensitive mashuped TV broadcasts, videos, additional information, news, services like weather or targeted travel offers as well contributed content and recommendations from the user in the channel community. Moreover the channel is connected to an interactive language learning system. The application is realized on the base of in this paper stated approaches for a modern TV2.0 based T-Learning environment and their possible pedagogical effects. The second project is not specialized on a certain subject. It focuses on immediate access to context sensitive additional information about the current showed topics and objects in a broadcast or video as well on the will for execution on immediately transactions for services and products related to the viewed content. Both projects are currently in the finishing as look&feel scenarios, which we want to investigate regarding the impacts of interaction, navigation and transaction.

4. REFERENCES

- [1] ARD/ZDF-Onlinestudie (2008). Zunehmender Medienkonsum: Internetnutzung bei 58 Minuten täglich. <http://www.ard-zdf-onlinestudie.de/index.php?id=125>
- [2] Atwere, D. & Bates, P. (2003): Interactive TV – a learning platform with potential. London, Learning and Skills Development Agency.
- [3] Baumgartner, P. (2008). Die zukünftige Bedeutung von Online-Lernen für lebenslanges Lernen. In: Issing, L.J. & Klimsa (Ed.), P. Online-Lernen – Handbuch für Wissenschaft und Praxis, (pp.505-514). München, Oldenbourg.
- [4] BBC (2008). Annual Report and Accounts 2007/08 The BBC Executive’s review and assessment. Online-Quelle: http://downloads.bbc.co.uk/annualreport/pdf/bbc_ara_2008_exec.pdf. Recherche: Juni 2008.
- [5] BITKOM (2006). Digitale Konvergenz – Empfehlungen. BITKOM Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., Berlin
- [6] Broadband World Forum Europe 2008: Multimedia Corner/Heine. http://www.iec.org/events/2008/bbwf/multimedia_corner.asp
- [7] Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods. New York: Irvington.
- [8] Cross, J. (2006). Informal Learning. John Wiley & Sons, Incorporated.
- [9] Dohmen, G. (2001). Das informelle Lernen. Die internationale Erschließung einer bisher vernachlässigten Grundform menschlichen Lernens für das lebenslange Lernen aller. Bonn: BMBF
- [10] Downes, S. (2007). E-Learning 2.0 - in development, <http://www.slideshare.net/Downes/elearning/-20-in-development>.
- [11] Duffy, T.M. & Cunningham, D.J. (1996). Constructivism: Implications for the design and delivery of instruction. In: Jonassen, D.H., Handbook of Research for Educational Communications-Technology, (pp.170-198). New York: Simon & Schuster MacMillan.
- [12] Ehlers, U.-D. (2008). Qualität für neue Lernkulturen des “Next Generation” E-Learning. In: Issing, L.J. & Klimsa (Ed.), P. Online-Lernen – Handbuch für Wissenschaft und Praxis, (pp. 339-356) München, Oldenbourg.
- [13] Fleming, M., & Levie, W. H. (Eds.). (1993). Instructional message design: Principles from the behavioral sciences (2d ed.). Englewood Cliffs, NJ: Educational Technology.
- [14] Forrester Research (2008). A Deep Dive into European’s Online Behaviour.
- [15] Gerhards, E., Klingler, W. (2008): Fernseh- bzw. Bewegtbildnutzung 2007. In: Media Perspektiven, Heft 11/2008, S. 550 – 567
- [16] Gleich, U. (2008): Muster und Funktionen der Medien-nutzung. In: Media Perspektiven, Heft 3/2008, S. 148 – 154.
- [17] Hasebrook, J. (2008). Online-Lernen in Banken. In: Issing, L.J. & Klimsa, P. (Eds.), Online-Lernen – Handbuch für Wissenschaft und Praxis, (pp.457-470) München, Oldenbourg.
- [18] Issing, L.J. & Klimsa, P., (Eds.), (2008). Online-Lernen – Handbuch für Wissenschaft und Praxis. München, Oldenbourg.
- [19] Jonassen, D.H. (2004). Handbook of research on educational communications and technology. Association for Educational Communications and Technology. Edition: 2, Lawrence Erlbaum Associates.
- [20] Jonassen, D.H., (2000). Revisiting Activity Theory as a Framework for Designing Student-Centered Learning Environments. In: Jonassen, D.H & Land, S.M. (Ed.), Theoretical Foundations of Learning Environments. (pp. 89-122). Lawrence Erlbaum Associates.
- [21] Kerres, M. (2003). Wirkungen und Wirksamkeit neuer Medien in der Bildung. In: Keil-Slawik, R. & Kerres, M., Wirkungen und Wirksamkeit neuer Medien in der Bildung. (pp. 31-44) Waxmann Verlag.
- [22] Microsoft (2009). EUROPE LOGS ON. European Internet Trends of Today and Tomorrow. <http://download.microsoft.com/download/9/5/1/951300C9-7F7C-4656-B00F-B19A72363669/EuropeLogsOn.PDF>
- [23] Moore, D.M., Burton, J.K., Myers, R.J. (1996). Multiple-Channel Communications: The Theoretical and Research Foundations of Multimedia. In: Jonassen, D.H. Handbook of Research for Educational Communications and Technology. Second Edition, (pp. 851-878). New York: Simon & Schuster MacMillan.
- [24] Moore, D.M., Dwyer, F.M., (1994). Visual Literacy: A Spectrum of Visual Learning. Educational Technology Publications.
- [25] Motorola (2008). Motorola Survey: Media Mobility is Key for the Millennial Generation in Europe & the Middle East. www.motorola.com/mot/doc/7/7176_MotDoc.pdf
- [26] Oehmichen, E. & Schröter, C. (2007): Zur typologischen Struktur medienübergreifender Nutzungsmuster. In: Media Perspektiven, Heft 8/2007, S. 406 – 421
- [27] Online Marketing Trends (2008). Web2.0 Trends. <http://onlinemarketingtrends.blogspot.com/search/label/web2.0%20trends>
- [28] Paivi Aarenniemi-Jokipielto 16th (2005). T-learning Model for Learning via Digital TV. Proceedings of the EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering (EIE)Lappeenranta, Finland6th-8th June 2005
- [29] Paolucci, R. (1998). The effects of cognitive style and knowledge structure on performance using a hypermedia learning system. Journal of Educational Multimedia and Hypermedia, 5, (pp. 239-281).
- [30] Prario, B. & Imberti Dosi, A. (2004). New Frontiers of T-Learning: the Emergence of Interactive Digital Broadcasting Learning Services in Europe. In L. Cantoni & C. McLoughlin (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004 (pp. 4831-4836). Chesapeake, VA: AACE.
- [31] PricewaterhouseCoopers (2008): IPTV – das neue Fernsehen? Online-Source: <http://www.pwc.de>
- [32] Roland Berger Strategy Consultants (2008): WebTV insights and perspectives- A web 2.0 phenomenon is coining new TV usage patterns.
- [33] Salomon, G. (1979). Interaction of media cognition and learning. San Francisco: Jossey-Bess
- [34] Schank, R.C. (1994). Active Learning through Multimedia. In: IEEE, Spring, (pp. 69-78)
- [35] Seels, B., Fullerton, K., Berry, L., Horn, L.J. (2004). Research on Learning from Television- In: Jonassen, D.H. Handbook of Research for Educational Communications and Technology. Second Edition, (pp. 249-334.) New York: Simon & Schuster MacMillan.
- [36] seomoz.org (2008). <http://www.seomoz.org/blog/advanced-web-20-at-the-shoporg-summit>
- [37] SevenOne Media (Hrsg.), (2007): Erlebnis Fernsehen. Alltagsmedien im Vergleich. Unterföhring: SevenOne Media
- [38] Siemens, G. (2004). Connectivism: A Learning Theory for the Digital Age <http://www.connectivism.ca>
- [39] Svoen, B. (2007). Educational Programmes in Television and New Media: Diverging Patterns. In C. Montgomerie & J. Seale (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, (pp. 542-547). Chesapeake, VA: AACE.
- [40] Telecommunications Online (2008). Broadband boom borders 120 million in Europe. http://www.telecommagazine.com/article.asp?HH_ID=AR_4498
- [41] YouTube (2009). YouTube - Über YouTube. http://www.youtube.com/t/fact_sheet
- [42] Zubayr, C. & Gerhard, H. (2008): Tendenzen im Zuschauerverhalten. In: Media Perspektiven, Heft 3/2008, S. 106 – 119.

TUTORIALS

'How to create an iDTV campaign' tutorial

Koen Swings, CTO

Zappware

Media 4

3500 Hasselt (B-)

0032(0)497406440

koen.swings@zappware.com

ABSTRACT

In this paper, we describe the outlines of the Zappware tutorial on 'How to create an iDTV campaign'.

Keywords

Interactive Advertising, interactive Digital TV, iFormats



1. A BRIEF DESCRIPTION

Participants get a short introduction after which they will be asked to create an iDTV campaign.

2. A DETAILED OUTLINE

1 hour Introduction

During the introduction Zappware will

- showcase some iDTV campaigns,
- describe the Zappware iFormats and
- demonstrate how to create an iDTV campaign using Zappware's Iview Platform.

1h30 workshop

After the introduction teams of maximum 4 participants will be asked to create an iDTV campaign themselves. In order to do so they will first have to brainstorm and come up with a concept. Then they will be given access to Zappware's platform for the creation of iDTV services. They will be able to choose a suitable iFormat after which they can start to create the iDTV campaign themselves. They will also have to create suitable images and write appealing texts. Two Zappware collaborators will be present to assist if necessary.

0h30 evaluation

Each team presents his own iDTV campaign. All the iDTV campaigns are evaluated by the participants and by the presenter and his two collaborators. The best iDTV campaign wins a bottle of Freixenet champagne.

3. CHARACTERIZATION

Introductory

4. LENGTH

Half-day

5. POTENTIAL TARGET AUDIENCE

Advertising agencies, namely their Account Directors, Account Managers, Content Managers but also TV stations, namely their IT Managers, Webmasters, Sales Managers, Account Managers.

6. PRESENTER'S RESUME

Koen Swings received a Phd in engineering from the University of Leuven in 1995, after incorporating new artificial intelligence techniques in the computer-aided design of integrated circuits.

He then moved to Philips where he spent most of his time managing an innovative software development division specialised in advanced software architecture for next-level digital tv set-top boxes.

In 2001 he founded Zappware, a Belgium-based company specialised in the creation, implementation and operation of interactive digital television services.

Koen Swings is the author of many papers on various topics, ranging from artificial intelligence to digital television, and has presented them at numerous conferences and workshops.



User Experience in TV-centric Services: What to consider in the Design and Evaluation?

Marianna Obrist
ICT&S Center
University of Salzburg
Sigmund-Haffner-Gasse 18
5020 Salzburg, Austria
marianna.obrist@sbg.ac.at

Hendrik Knoche
Human-Centred Systems
Dep. of Computer Science
University College London
Gower Street, London WC1E 6BT, UK
h.knoche@cs.ucl.ac.uk

Damien Alliez
NDS Technologies France
65, rue Camille Desmoulins
94442 Issy-les-Moulineaux Cedex
France
dalliez@nds.com

ABSTRACT

The purpose of this tutorial is to motivate researchers and practitioners to think about the challenging question how to design for and evaluate user experience in TV-centric services. There is no common understanding of user experience yet, but a lot of different approaches and definitions (e.g. [1][2][3][4]). In particular, when trying to understand and address user experience in the design and development process of new TV-centric systems or services, clear guidelines are often missing. Within this tutorial, we provide not only a state-of-the-art overview on user experience but want to make a first step towards a better understanding on this concept in the EuroTV community.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Documentation, Human Factors.

Keywords

Implications for Design, Evaluation, TV-centric Services

1. INTRODUCTION

The goal of this tutorial is to give the participants an overview on lessons learned and best practices in the design and evaluation of user experience in TV-centric services. The focus will be to bring concrete examples and to outline some main implications for design and evaluation, such as:

When and how to involve users? How to avoid user overload? How to translate user feedback into requirements for designers? How to obtain the best results on conceptual designs and prototypes from respective target groups?

Our main focus is on user experience, as a major goal of design and development activities in research and industry. Usability is no longer the only aspect to be considered, but also the user experience, which consists of several factors, such as fun, motivation, emotion and sociability. At the end of this half-day tutorial, the participants should be able to draw major conclusions and understand the implications to consider in the design and evaluation of TV-centric services (possibly on TV, PC or mobile) in a user-centred design process.

As tutorials should follow an interactive and participative approach, we will provide the participants with some initial input

founded on our expertise on the above questions, followed by an interactive discussion on the participants' thoughts. Based on selected use cases or systems we will prepare some provocative statements on user experience design and evaluation to elicit views and responses from the participants. For example, NDS will present the Unified EPG (Oasis) system and the new Snowflake system, designed and evaluated towards user experience aspects in laboratory environments as well as in the field. A live presentation of Snowflake, which combines TV, PC and mobile into one platform, is possible on site. Thereby the tutorial ensures a practical focus based on a common theoretical and empirical basis.

2. AUDIENCE

The tutorial aims to bring together researchers on User Experience from a scientific and industrial background. To ensure the reachability of these two diverse audiences we combine scientific and industrial competence and expertise in the organizing team of this tutorial.

Target participants are in particular designers of TV-centric services, as well as researcher concerned with the evaluation of such services. Their major goal is to provide users with the optimal experience when engaging with an interactive system. The tutorial aims to attract both experienced designers and researchers around the field on User Experience, but also affords newcomers a clear overview at what is going on in research and industry with regard to design and evaluation of interactive TV-centric systems.

3. SCHEDULE

The tutorial is foreseen as a half-day event, combining theoretical and empirical parts as well as interactive hands-on exercises. In the first part of the tutorial we will present a comprehensive background on the topic and stimulate an interactive discussion. The participants will have the opportunity to apply and internalize this knowledge and share their own expertise in a hands-on exercise in the second part of the tutorial. The overall goal of this combination is to learn from each others' experiences based on clearly defined topics and guiding questions.

09:00 – 09:30: What is and Why to care about User Experience?

- Understand User Experience
- Design for User Experience
- Evaluate User Experience

09:30 – 10:00: TV-centric Services – Practical Examples

- Lessons learned from UNIC project
- Lessons learned from CITIZEN MEDIA project
- Implications for Design & Evaluation

10:00 – 10:30: Coffee Break

10:30 – 12:00: Interactive Session

- Hands-on Experiences with selected Examples
- Discussions in Sub Groups based on pre-defined Topics, such as: Design for User Experience? Evaluate User Experience? Constraints and Challenges for TV-centric services? New methodological approaches (playful and creative methods and techniques)

12:00 – 13:00: Plenary Discussion & Conclusions

- The most important conclusions from the previous discussions in the sub groups (5 bullet points) should be presented and discussed in a final plenary.
- Overall Conclusions summarized by the organizers with regard „What to consider in the Design and Evaluation of TV-centric Services?“

The schedule provides a framework for gathering a better understanding of user experience research around TV-centric services and a forum to elaborate on and share individual knowledge with other participants from research and industry.

4. INSTRUCTORS

Marianna Obrist is Assistant Professor in the HCI & Usability Unit of the ICT&S Center at the University of Salzburg. She holds a doctoral degree in Communication Science from the University of Salzburg. The focal point of her research lies in user-centered design methods, and in particular the user involvement into the development of new products/systems. She is involved in several research projects concerned with the study of user requirements and the evaluation of user experience and acceptance for iTV, IPTV (i.e. iiTV@home, iTV 4 all, CITIZEN MEDIA). She has co-organized a workshop at CHI'07 and a Special Interest Group at CHI'08, and is recently co-organizing a workshop on social media at the NordiCHI08 conference. She was part of the organization team for the MobileHCI 2005, ACE 2007, and is recently conference co-chair for EuroITV 2008.

Hendrik Knoche is finishing his PhD at University College London in computer science in the Human Centred Systems Group. He holds a Diploma in computer science from the University of Hamburg. His latest research and thesis focus on Quality of Experience (QoE) in mobile multimedia services. He has looked into beneficial trade-offs between QoE and economic constraints to deliver satisfying mobile TV experiences. Hendrik has worked in industry as a consultant in information architecture, mobile services (Nokia, Vodafone, Sport1) and research projects investigating distributed collaboration and TV-centric services. He is on the program committee of EuroITV, uxTV and the CHI'09 workshop on mobile user experience and recently organized a workshop on user experience in TV-centric services.

Damien Alliez is Engineer with Multimedia Network Specialization degree. He has been working as consultant for Lucent Technologies (EMEA CTI products marketing manager) in 1998, for GEMPLUS (Telco and Network infrastructure definition and installation) in 1999 and for CANAL+ Technologies (System engineering, design and development (HP-UX), installation and integration of a Mail solution in a terrestrial Pay TV environment) in 2000. He joined NDS France in 2000 as technical project manager. Specialized in Digital TV Head-End architecture (satellite, terrestrial, cable, IPTV) and after managing several operational projects, he is now responsible for the collaborative research projects and his company representative in several standard bodies like DVB and ITU-T. He is currently chairing the DVB IPI Home Network group

5. Related Experiences & Publications

The instructors of this tutorial have been actively involved in the previous EuroITV conferences, as organizers, presenters and exhibitors. Some selected publications are listed below outlining the instructors' field specific experience:

- **Alliez, D.** (2008). Adapt TV Paradigms to UGC by improving social networks. In Adjunct Proc EuroITV 2008, 29-30.
- Boccolini, G., **Knoche, H.**, Luise, M., Merour, J.-M., Vignaroli, L. (2008) Broadband interactive services provided through TV-centric technologies via bidirectional satellite connection. In Adjunct Proc. EuroITV 2008.
- **Knoche, H.**, Sasse, M. A. (2008) Getting the big picture on small screens: Quality of Experience in mobile TV. In Ahmad, A. M. A. & Ibrahim, I.K. (eds.) Multimedia Transcoding in Mobile and Wireless Networks, Chapter 3, pp. 31-46, Information Science Reference.
- **Obrist, M.**, Bernhaupt, R. & Tscheligi, M. (2008). Interactive TV for the Home: An Ethnographic Study on Users' Requirements and Experiences. In International Journal of HCI, Vol 24, 2 (2008), 174-196.
- **Obrist, M.**, Bernhaupt, R., Beck, E., & Tscheligi, M. (2007). Focusing on Elderly: an iTV Usability Evaluation Study with Eye-Tracking. In: Cesar, P., Chorianopoulos, K., Jensen, J. F. (Eds.). Interactive TV: A shared Experience. In Proc.EuroITV 2007, 66-75.

6. REFERENCES (Selected)

- [1] Forlizzi, J. & Battarbee, K. (2004). Understanding experience in interactive systems. In Proc Designing Interactive Systems 2004, NY: ACM, 261.
- [2] Hassenzahl, M., & Tractinsky, N., (2006). User Experience - a research agenda. In Behavior & Information Technology, 25(2), pp. 91-97.
- [3] Karapanos, E., Hassenzahl, M., Martens, J.-B., (2008). User Experience Over Time, Work in Progress, In: Proc CHI 2008, ACM, New York, USA, pp. 3561-3566.
- [4] Law, E., Roto, V., Vermeeren, A., Kort, J., Hassenzahl, M. (2008). Towards a shared definition of user experience. In Proc CHI 2008, 2395-2398

Ambient Media – An Introduction by Case-Studies

EuroITV 2009 Tutorial

Artur Lugmayr
Tampere Univ. of Technology (TUT)
NAMU Lab.
POB. 553, Korkeakoulunkatu 1
Tel. +358 40 821 0558

lartur@acm.org
www.cs.tut.fi/~lartur

ABSTRACT

Media evolved from media that can be described as integrated presentation in one form (multimedia). From multimedia, media evolved towards embedding the consumer in a computer graphic generated synthetic world (virtual reality). From this point on, media evolved to the consumers directly exposed to the media in their natural environment, rather than computer interfaces (ambient media). In addition, media will be evolving towards a fully real/synthetic world undistinguishable from pure media integrating human capacity (biomedia or bio-multimedia) somewhere in the very far distant future. The goal is to train and educate participants in new innovative service design for ambient multimedia. The course will cover potential and possibilities of this new multimedia field and its relation to other trends, such as ubicom, pervasive computation, affective computation, and tangible media. Specific key-concepts of ambient media are developed based on various business case studies.

Keywords

Ambient media, ambient computation, ubiquitous computation, pervasive computation, smart media, Web 2.0, context awareness, emotional computation.

1. OBJECTIVES

The objective of this tutorial is that the audience is able to answer the following questions:

What are ambient media? What is the form and language of ambient media? What is ambient intelligence and which technologies does it bring?

The audience shall be trained in the emerging media environment and know its basic concepts. For the HCI community, it presents novel methods for interaction and navigation through media in the consumer surrounding environment. For the media interested communities, it gives a glimpse what changes ambient media brings for existing media eco-systems. For the more technical focused audience, technical systems and ambient computation will be introduced. In addition to previously held tutorials, this tutorial presents 5-10 case studies on general level. To deepen the understanding of ambient media, the tutorial examines selected case studies under technological, HCI, consumer experience, and smart media aspects.

2. CONTENT OF THE TUTORIAL

2.1 Part 1: Introduction

The introduction gives insights into the development of ambient media. It reviews current trends and gives details into the

development of ambient media in general. In further detail, the introduction contains the following main issues:

- History, general viewpoints, and business aspects
- State-of-the art, standardization efforts, and organizations
- Legal factors, socio-economic, and privacy aspects
- Concepts overview, trends, and technology enables
- Ambient media form, content, and techniques
- First definition of ambient media

2.2 Part 2: Case Studies

The second part of the tutorial introduces 5-10 case studies of ambient media on general level. These case studies will be refined within the scope of the media eco-systems part of the tutorial, and the technical part of the tutorial.

- Business case studies from the media management, marketing, strategically, and service point of view
- Business case studies for ambient media
- Presentation of Nokia Ubimedia MindTrek Award winning projects from 2007 and 2008 (see www.mindtrek.org)
- Case studies form the SAME 2008 and other ambient media related workshops held in conjunction with other events

2.3 Part 3: Smart Interactive Media Eco-Systems

Definition and exploitation of ambient media from a media and human-computer-interaction viewpoint. This part of the tutorial shall answer the following questions: what are ambient media? What is the form of ambient media? How will HCI change and how can we design for ambient? The following are the main issues to be discussed within the scope of this part of the tutorial:

- Principles of ambient media
- HCI and ambient media
- Natural environment as place for media
- Interaction models and behavior patterns
- Consumer experience and consumer studies
- Context awareness
- Personalization and consumer profiling
- Emotions and media
- *Examination of case-studies on these aspects*

2.4 Part 4: Ambient Media Systems

Ambient intelligence is the basic technology enabling ambient media. Within this part of the tutorial the system aspects shall be evaluated:

- Ambient and cross-service oriented computation
- Hardware components

- Smart materials
- MEMS & sensor technology
- Adaptive software
- Ubiquitous communication
- I/O device technology
- Software components
 - Media management & handling
 - Emotional computation
 - Natural interaction
 - Context awareness
 - Computational intelligence
 - Presence technologies
- *Examination of case-studies on these aspects*

3. TARGET AUDIENCE

This course is designed for a general audience, which is interested in the future of media technology. Its content is relevant for an audience with industrial, media, technological, or service design interest.

4. SCHEDULE OF THE TUTORIAL

The tutorial is scheduled for Wednesday afternoon, 3rd of June 2009, 14:00-18:00 (4 hours):

14:00-15:00 (1h) : Part 1: Introduction

15:00-15:30 (1/2h): Part 2: Case Studies

15:30-16:00 : Coffee Break

16:00-17:00 (1h) : Part 3: Smart Interactive Media Eco-Systems

17:00-18:00 (1h) : Part 4: Ambient Media Systems

5. READINGS & MATERIAL

For further reading, please check out the *Ambient Media Association* (AMEA) website at: <http://www.ambientmediaassociation.org>. Much of the material can be found on the association's website, where also more material can be downloaded. Other interesting readings are: [1-14]

REFERENCES

- [1] A. Lugmayr, "The future is 'ambient'," in *Proceedings of SPIE Vol. 6074, 607403 Multimedia on Mobile Devices II*. vol. 6074, R. Creutzburg, J. H. Takala, and C. W. Chen, Eds. San Jose: SPIE, 2006.
- [2] A. Lugmayr, "Ambience, ambience, ambience - What are Ambient Media?," in *Interactive TV: A Shared Experience, TISCP Adjunct Proceedings of EuroITV 2007*, Amsterdam, 2007, http://www.cs.tut.fi/~lartur/euroitv07_ajp/main.htm.
- [3] A. Lugmayr, "Ambient Media," *Novatica*, vol. 33, pp. 35-39, July-August 2007, http://www.ati.es/novatica/infonovatica_eng.html.
- [4] A. Lugmayr, "From Ambient Multimedia to Bio-Multimedia," in *High Performance Multimedia - A Reader on the Technological, Cultural and Economic Dynamics of Multimedia*, P. A. Bruck and J. Boumans, Eds. Amsterdam: IOS Press, 2008.
- [5] A. Lugmayr, *Ambient Media and Beyond*: Springer-Verlag NY, 2008 (submitted), <http://www.springer.com/west/home/computer/communications?SGWID=4-148-22-173707313-0>.
- [6] A. Lugmayr, F. Mäyrä, K. Lietsala, and H. Franssila, "Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era," Tampere, Finland, 2008, p. 221.
- [7] A. Lugmayr, A. Pohl, M. Mühlhäuser, J. Kallenbach, and K. Chorianopoulos, "Ambient Media and Home Entertainment," in *Interactive Digital Television: Technologies and Applications*, G. Lekakos, K. Chorianopoulos, and G. Doukidis, Eds.: Idea Group Publishing, 2007.
- [8] A. Lugmayr, S. Reymann, S. Kemper, T. Dorsch, and P. Roman, "Bits of Personality Everywhere: Implicit User-Generated Content in the Age of Ambient Media," in *Parallel and Distributed Processing with Applications, 2008. ISPA '08. International Symposium on*, 2008, pp. 516-521.
- [9] A. Lugmayr, B. Stockleben, T. Risse, J. Kaario, and K. Laurila, "ACM multimedia 2008: 1st workshop on semantic ambient media experiences (SAME2008) namu series," in *Proceeding of the 16th ACM international conference on Multimedia* Vancouver, British Columbia, Canada: ACM, 2008.
- [10] A. Lugmayr, B. Stockleben, T. Risse, J. Kaario, and K. Laurila, "Semantic Ambient Media Experiences SAME 2008 Pre-Workshop Review (NAMU Series).," in *Proceeding of the 16th ACM international conference on Multimedia (SAME 2008 workshop)* Vancouver, British Columbia, Canada: ACM, 2008.
- [11] A. Lugmayr, A. Pohl, M. Mühlhäuser, D. Negru, J. Kallenbach, F. Köbler, and S. Niiranen, "Ambient Media and Home Entertainment (Workshop Review)," in *Proceedings of the 4th Euro iTV Conference* Athens, Greece: ELTRUN/AUEB, 2006.
- [12] S. Reymann, S. Kemper, and A. Lugmayr, "AmbiNet - A Lightweight Open Source Development Framework for Personalized Ambient Entertainment," in *Changing Television Environments, TICSP Adjunct Proceedings of EuroITV 2008*. vol. 42, A. Lugmayr, S. Kemper, M. Obrist, T. Mirlacher, and M. Tscheligi, Eds. Salzburg, Austria: Tampere University of Technology (TUT), 2008.
- [13] S. Reymann, J. Rachwalski, S. Kemper, and A. Lugmayr, "Development of a Generic XML Personality Metadata Handler for Distributed Entertainment Services," in *EuroITV (Changing Television Environments)*, 2008, <http://www.springerlink.com/content/v227275r18h61718>.
- [14] R. Simon, S. A. David, and L. Artur, "Personalized social networking: an applied scenario in a portable personality environment," in *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era* Tampere, Finland: ACM, 2008.

Tutorial: Designing and Evaluating the Sociability of Interactive Television

David Geerts

CUO, IBBT / K.U.Leuven

Parkstraat 45 Bus 3605

3000 Leuven, Belgium

+32 16 32 31 95

david.geerts@soc.kuleuven.be

ABSTRACT

In this tutorial, we will discuss how the social uses of television have an impact on how we should design and evaluate interactive television applications. We will focus on the concept of sociability, and explain how this can be evaluated by using sociability heuristics. We will also discuss how sociability can be studied by performing user tests, and which aspects of testing are different from studying usability. Although the guidelines and user tests are especially appropriate for designing and evaluating social television systems, parts of it are also suitable for other iTV applications.

Categories and Subject Descriptors

H.4.3 Information Systems Applications: Communications Applications, H.5.2 User Interfaces: Evaluation/methodology, H.5.1 Multimedia Information Systems: interactive television

General Terms

Measurement, Design, Experimentation

Keywords

Sociability, heuristics, evaluation, social television

1. INTRODUCTION

In past EuroITV conferences, as well as at uxTV2008, social TV has proven to be an important and exciting new topic of research in interactive television. Many social TV systems are currently being developed in academic as well as corporate research labs [1, 2, 4, 6, 7]. As is good practice in user-centered design, evaluating these systems early and often is important to create an optimal user experience. Although several guidelines for evaluating the usability of interactive TV exist [3, 9], and heuristic evaluation [11] as well as usability testing is a well-known and often practiced technique, for applications being used in a social context such as the social television systems mentioned above, evaluating only usability is not enough. Even if these applications are evaluated to improve their usability, it doesn't mean that the social interactions they are supposed to enable are well supported.

This tutorial wants to fill this gap by teaching researchers and practitioners how to design and evaluate social features of interactive television. Based on his extensive experience in performing user tests of social television systems for evaluating their sociability (e.g. [2, 5]), the presenter will explain the practical issues related to performing user tests with iTV focused on social interactions. Furthermore, he will discuss the sociability heuristics he has created based on these tests and explain how they can be used to evaluate social television systems or social aspects of interactive television in general.

The proposed structure of the tutorial is as follows: first, the social uses of television as documented by several media researchers [8, 10] will be introduced. They will be linked with the current state of interactive television services and applications, including a wide range of social TV systems, and the need for designing and evaluating sociability will be explained. Then, a detailed overview of twelve sociability heuristics the presenter has developed will be given, along with an explanation of how to use them to evaluate iTV. After this, practical issues when performing user tests focused on social interaction will be discussed, including tips of how to improve them. Finally, a practical hands-on session will be held in which the participants can apply the sociability heuristics to an online social television application such as Lycos Cinema (or another suitable application that will be available at the time).

2. SCHEDULE

- 1/2 hour introduction to the social uses of (interactive) television
- 1 hour overview of sociability heuristics
- 1/2 hour discussing sociability user testing
- 1 hour practical exercise

3. TARGET AUDIENCE

The target audience for this course is researchers and practitioners that specialize in designing or evaluating

interactive television applications, and that want to increase their focus on social aspects of iTV. The tutorial requires basic knowledge of the different kinds of interactive television applications as well as of usability (either general or specific for iTV). No knowledge of social TV or sociability is required.

4. LEVEL OF THE TUTORIAL

The level of the tutorial is intermediate, as some knowledge on iTV applications and usability is required.

5. PRESENTER'S BIO

David Geerts has a master in Communication Science at the K.U.Leuven and a master in Culture and Communication at the K.U.Brussel. He was project leader of the Mediacentrum of the Katholieke Universiteit Leuven (Belgium) for several years, and now leads the Centre for User Experience Research (CUO). He is involved in several research projects on user-centered design and evaluation. Furthermore, he acts as content manager for the post academic course Usability Design. David has over ten years of experience in teaching for a diverse audience: e.g. introductions to new technologies for secondary school children, practical seminars in web design for university students and in-depth courses on usability topics for practitioners. Currently he teaches a master course in Human-Computer Interaction for communication scientists, economists and industrial engineers, as well as usability design courses and workshops for practitioners. David Geerts has organized workshops and SIGs at CHI2006, CHI2007, CHI2008, EuroITV2007 and EuroITV2008. He finished his doctor's degree on Sociability of Interactive Television, for which he has developed twelve heuristics for designing and evaluating social television interfaces. David Geerts is co-founder and board member of the Belgian SIGCHI.be chapter, and is program chair of EuroITV2009, the 7th European Interactive TV Conference.

6. REFERENCES

- [1] Boertjes, E., Schultz, S. and Klok, J. Pilot ConnecTV. Gebruikersonderzoek. Public report of the Freeband Project, TNO (2008)
- [2] Cesar, P., Bulterman, D.C.A., Geerts, D., Jansen, A.J., Knoche, H., and Seager, W.,. Enhancing Social Sharing of Videos: Fragment, Annotate, Enrich, and Share. In Proc. of ACM MM '08, ACM (2008)
- [3] Chorianopoulos, K. User Interface Design Principles for Interactive Television Applications. In International Journal of Human-Computer Interaction, 24, 6, Taylor Francis (2008)
- [4] Coppens, T., Trappeniers, L. & Godon, M. AmigoTV: towards a social TV experience. In Masthoff, J. Griffiths, R. & Pemberton, L. (Eds.) Proc. EuroITV2003, University of Brighton, (2004)
- [5] Geerts, D. 2006. Comparing voice chat and text chat in a communication tool for interactive television. In Proc. NordiCHI 2006, 461–464.
- [6] Harboe, G., Massey, N., Metcalf, C., Wheatley, D., & Romano, G. The uses of social television. In ACM Computers in Entertainment (CIE), 6, 1 (2008)
- [7] Harrison, C., Amento, B., 2007. CollaboraTV: Using asynchronous communication to make TV social again, In Adjunct Proceedings of EuroITV2007, 218-222.
- [8] Lee, B. & Lee, R.S. How and why people watch tv: implications for the future of interactive television. In: Journal of advertising research, 35, 6 (1995), 9-18
- [9] Lu, Karyn Y. Interaction Design Principles for Interactive Television. Master's Thesis, Georgia Institute of Technology (2005)
- [10] Lull, J. The social uses of television. In: Human Communication Research, 6, 3, Spring, (1980) 197-209
- [11] Nielsen, J. Heuristic evaluation. In Nielsen, J., & Mack, R. L. (Eds.), Usability inspection methods. John Wiley & Sons (1994), 25-62

Tutorial EOL for EuroITV 2009

Standardized IPTV services in Managed IP Infrastructures: An insight on approaches from the Telco & CE industry with a focus on NGN & IMS

Oliver Friedrich
Fraunhofer Institute FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin
+49 30 3463 7145

oliver.friedrich@fokus.fraunhofer.de

Robert Seeliger
Fraunhofer Institute FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin
+49 30 3463 7145

robert.seeliger@fokus.fraunhofer.de

Keywords

Interactivity, IPTV, NGN, Prototype, Standardization

1. Tutorial Description

This document provides information on a half day tutorial proposal on standardized IPTV solutions to be held at EuroITV 2009 conference in Leuven, Belgium.

1.1 Tutorial Characterization

This tutorial is classified as intermediate or advanced with regards to the Tracks on Technical Insights. Introductions, Track 1, 2 and 4 including Live Demonstration and usability Track are not requiring any knowledge about the topic.

1.2 Target Audience

Information given in this tutorial will address anyone interested in getting a deep insight into IPTV standardization, service signaling, service creation, prototyping and usability. Prerequisite telecommunications knowledge is not mandatory but will help to understand some of the aspects more deeply.

2. Detailed Tutorial Outline

This half day tutorial will give a deep inside on IPTV over telco driven Next Generation Networks and other concurrent approaches from the Consumer Electronics Industry.

In detail this includes ongoing IPTV standardization activities under the umbrella of Next Generation Networks (NGN) and the IP Multimedia Subsystem (IMS) at involved Standard Development Organizations (SDOs) as ETSI TISPAN, Open IPTV Forum, ITU-T IPTV GSI and DVB. Furthermore the different views of Telco operators & industry, Consumer Electronics (CEs) and content providers will be analyzed which includes a discussion of so called Telco walled garden vs. Portal walled gardens or "Who owns the customer" issues. In a third track available services, service signaling & user experience issues will be presented followed by a fourth track introducing a real world test bed for converged media services called FOKUS MI Lab (www.mediainteroperabilitylab.org) using the well known Fraunhofer FOKUS Open IMS Playground as signaling

infrastructure. A live demonstration and hands on session might be included in this tutorial.

Track 1: An introduction to IP based streaming services & IPTV

This track will introduce and analyze the basic IPTV scenarios and requirements towards IPTV infrastructures and will try to find a definition for IPTV and corresponding services, e.g.:

- IPTV vs. Web TV
- Managed vs. unmanaged infrastructures
- Basic Scenarios & Requirements

Track 2: Standardization & Stakeholder Analysis

This track will give a deep overview on the different Standard Development Organizations (SDOs) working on standards for IPTV. In additions this track will analyze the different views and strategies of the key players in the IPTV market. This includes:

- Telecom Operators
- Broadcasters
- Content Providers
- Consumer Electronics
- Advertisement Industry

Track 3: Technical insights on NGN based IPTV systems

This track addresses details on NGN based conversational and media services evolving towards all-IP, converged Rich Media Infrastructures. In detail this includes a presentation on:

- Architectural Approaches
- IPTV service signaling
- Interaction with telecommunication services
- Application Server Models

Track 4 Live Demonstration & Usability

In this track the Fraunhofer FOKUS Media Interoperability Lab will be presented by showing the ecosystem in live action. In addition participating conference guests are invited to take part in a short usability test having a hand on the ecosystem by using the different services and the different Remote Control devices available (e.g. RC, Wii-Remote, iPhone

Live Demonstration

Usability Experience

3. Tutorial Presenters

3.1 Oliver Friedrich

Oliver Friedrich is a senior research engineer at the Competence Center Smart Environments at the Fraunhofer Institute for Open Communication Systems (FOKUS). He holds a M.Sc. in Computer Engineering from the Technical University of Berlin, Germany with specific emphasis on telecommunication networks.

Currently he is leading the IPTV & Converged Media activities at the Fraunhofer FOKUS Media Interoperability Lab (MI Lab) managing the integration of new services and working on architectural issues. In addition he is preparing his PhD in the field of personalized and interactive IPTV services delivered over Next Generation Networks.

He is actively involved in the IPTV standardization process at ETSI TISPAN and published several papers and articles in this context.

3.2 Robert Seeliger

Robert Seeliger is research engineer at the Competence Center Smart Environments at the Fraunhofer Institute for Open Communication Systems (FOKUS). He received his M.Sc. in Multimedia and Communication Technology from University of Jena in 2006.

He is responsible for the technical maintenance and ongoing developments within the FOKUS Media Interoperability Lab.

Robert is involved in several projects with regards to IPTV dealing with the integration of these services onto vehicular environments as well as FOKUS' cooperation with partners from the media and advertisement industry.

4. Contacts

Oliver Friedrich & Robert Seeliger

Fraunhofer Institut für Offene kommunikationssysteme

Kaiserin-Augusta-Allee 31

10589 Berlin

Phone: +49 30 3463 7145

Fax: +49 30 3463 8145.

WORKSHOPS

Workshop 1:
Think Positive – designing and
understanding enjoyable interactive media
experiences

Think positive – designing and understanding enjoyable interactive media experiences

Marc Hassenzahl
Folkwang University
Campus University of Duisburg-Essen
Universitätsstraße 12, 45141 Essen
+49 (0) 201 183-4482

marc.hassenzahl@folkwang-
hochschule.de

Susanne Sperring
iDTV Lab, MediaCity
Åbo Akademi University
P.B. 311, 65101 Vaasa, Finland
+358 (0)6 324 7405

susanne.sperring@abo.fi

Annika Wiklund-Engblom
iDTV Lab, MediaCity
Åbo Akademi University
P.B. 311, 65101 Vaasa, Finland
+358 (0)50 322 3458

annika.wiklund-engblom@abo.fi

ABSTRACT

Vendors, producers, and designers of interactive products and content become increasingly aware and interested in the notion of *Experience*. Experience-oriented design is a particular approach, which differs from others with respect to its strong focus on the needs and emotions of people and the positive aspects of product use. This workshop's objective is to take an experience-oriented perspective on the current practices in the design of iTV, related cross- and interactive media. We collect, share and discuss design ideas and solutions, which claim to be successful in enhancing the Experience. We will answer the two central questions of "What is experience?" and "How do we create it?" rather by a joint analysis and reflection of existing projects, their strengths and weaknesses (i.e., bottom-up) than by referring to abstract theoretical and methodological concepts (top-down). By that, we hope for insights into experience-oriented design in the iTV domain. These insights will be grounded on the expertise of people working in the field and are brought about by sharing and a guided reflection on their practices and concepts.

1. INTRODUCTION

Recently *Experience* became a buzzword not only in the context of practical and academic Human-Computer Interaction (as "User Experience", see [1]), but also in environments such as electronic games (e.g., [4]), and interactive television (iTV, e.g., [3]). Vendors, producers, or designers become increasingly interested in understanding and adequately addressing the experiences people have while using or consuming their products (e.g., shows, games, software, music, services).

An experience-oriented approach to design differs from other approaches (e.g., usability engineering, technology-oriented approaches) by stressing the *holistic*, *subjective*, and *positive* [2].

Holistic. Experience-oriented design aims for an appropriate balance between functional/usability issues (task-related aspects) and issues, such as beauty, challenge, stimulation, or self-expression. It also understands basic human needs, values and motivations, such as a need for novelty and change or social exchange, as the starting point of any creation/design activity (as opposed to starting from available technology or tasks).

Subjective. Experience-oriented design embraces the "subjective." It is interested in the way, people perceive, experience, judge, feel about the products they use or consume. Those "subjective

constructions" matter for at least two reasons: first, they will guide the future behavior of an individual ("SO bad. I will never do this again!"); second, they will be communicated to others ("have you heard about the latest show"). In other words, it does not matter how good a product is "objectively" or from an expert point of view. Any "quality" must be "experienced" to have impact.

Positive. Many design-approaches focus on barriers, problems, frustration, stress, and their removal. While removal of the negative will always remain an important aspect of product design, experience-oriented approaches stress the importance of positive outcomes of product use or consumption, such as positive emotions (e.g., joy, pride, excitement) or simply "value". This stems from the insight that "positive" may not necessarily equate to "absence of the negative." So far, much attention has been devoted to the negative, that is, the usability of, for example, iTV products and services. Certainly, there is more about a positive iTV experience than the absence of usability problems.

This workshop's objective is to take an experience-oriented perspective on the current practices in the design of iTV, related cross- and interactive media. We aim at collecting, sharing, and discussing design ideas and solutions, which claim to be successful in enhancing the Experience. By that, we seek to understand better the experiences people have when consuming content; interacting with it, sharing it, and creating mashups. What benefits do people get from iTV and related new forms of media? What are the ends, the active media consumer wants to achieve? In addition, by identifying successful cases, we hope to make explicit some of the methods and underlying principles of good experience-oriented design in the iTV domain. In other words, we want to identify best practices and lessons learned, and how these can be generalized as guidelines for creating positive user experiences with iTV and related crossmedia solutions.

2. WORKSHOP

The workshop is primarily aimed at designers/producers/creators of iTV products (services, content, applications) either with an industrial or academic background. It takes a bottom-up approach to answer two central questions: "What is experience?" and "How do we create it?" Instead of invoking and imposing theories and models of experience-oriented design (top-down), we try to answer both questions by sharing and discussing examples of experience-oriented projects (bottom-up). By that, we hope to

create a common understanding of experience-oriented design for iTV. This understanding is supposed to be "better" due to its clear grounding in the practices, experiences, skills, and the knowledge of people working in the iTV domain.

Accordingly, participants submitted a brief *project description*, giving an overview of the product itself, why it is supposed to be a good example for experience-oriented design, its status (concept, prototype, product), the methods employed within the design process and potential indicators of the product's success. We set up a web page at www.ux-research.org providing the details of the workshop and further information. Participants were selected on the basis of a clear statement why their project qualifies as experience-oriented design.

The workshop itself consists of three parts:

What have you done? Each participant (or group) gives a brief overview of their project. This is meant as input for all further discussions.

What is Experience? Participants (small groups of max. 5) discuss the similarity and differences in their understanding of experience embedded in and expressed by their product. What kind of experience did they want to create, where does that particular understanding of experience come from? etc. Each small group is required to produce a conceptual "map" showing the similarities and differences in their understanding. These will be presented to the plenum and further integrated. The result is an overview of shared and alternative concepts of experience in the iTV domain.

What made the success? In a second round, participants (small groups of max. 5) will discuss their approaches (i.e., methods) to create a good experience. They will create a "map" of usual and unusual methods, experiences with this methods, etc. After presentation and further integration, the result is an overview of experience-oriented methods in the iTV domain.

Overall, the workshop is a guided reflection of concepts and methods used by the participants in their daily design work.

3. PARTICIPANTS AND PROJECTS

We accepted seven projects for the workshop to enable and fuel an intensive discussion. They are diverse and cover ground, ranging from interactive experimental movies to utilitarian media player software; from content to technology. In the following, we describe each project briefly (presented in alphabetical order of the author). More details are available at www.ux-research.org.

3.1 Vocomedia (Regina Bernhaupt)

Vocomedia is an interaction concept for the living room allowing users to simply "individualize" their iTV services. By using a six-key navigation combined with a fingerprint reader and a so called "look-there" user interface metaphor, Vocomedia lays a basis for an easy to use interactive TV service by at the same time providing a measurable positive user experience (in terms of "hedonic quality"). Vocomedia will be deployed in the near future by a European Telco/IPTV provider.

3.2 XDOMO (Massimo Deriu)

XDOMO aims at the integration of internet and television technologies (crossmedia) by linking television content with web

services and vice versa. The starting point is that integration of internet and television implies that digital objects may migrate between both worlds. This involves rethinking a digital television appliance in terms of a complex system of interconnected components, which cooperate to merge the features of the two worlds without any forced changes of their own respective structure.

3.3 Transparency (Carlos Caires)

Transparency is a cinematographic project to study the potential (and limits) of an interactive filmic narrative. To create a narrative, the spectator can combine a series of 20 autonomous scenes around the themes of treachery, jealousy, narcissism, pain, love and sex. Transparency is an installation, which features a particular input device (a transparent cube) to support the spectator in develop the story.

3.4 A player for video content in social media environments (Monique de Haas)

A user centred design-approach was used to build an embeddable video player for making available Dutch Public Broadcasting (DPB) Video Content in social media environments. The task was to create a player that "seduced" people to prefer it to already broadly available alternatives. We achieved this by embedding features deeply rooted in the particular requirements of social media (exchange, social search, recommendation).

3.5 Hotel Media-Center System (Jan Hess, Torben Wiedenhoefer)

A Hotel Media-Center System (HMS) offers a variety of services to guests of several hotel chains. Empirical studies and expert reviews to optimize the system. Enjoyment encompassed ease-of-use (e.g., ease of setting the alarm clock) as well as more hedonic, non-utilitarian aspects, such as animations or the aesthetics of sound feedback.

3.6 My experiencelapland.tv (Mikko Manninen, Anthony Okuogume, Ari Alm, Antti Haase, Eero Leppänen)

Mopaali is Kemi-Tornio University of Applied Sciences' flagship for developing web media, social media and cross media instruments. Experiencelapland.tv (eltv), an interactive web television channels, is one of the Mopaali's pilot applications. Eltv allows for prospective experience of a potential tour through Lapland through available videos. After a real tour, travellers are encouraged to upload own videos to document and share their experience.

3.7 The Space Trainees (Simon Staffans, Anders Wik)

The Space Trainees is a cross media edutainment about language learning targeted at children from eight to twelve years. It started out as a showcase for MHP-interactivity and developed into a cross media edutainment show. This was possible through thorough testing and evaluation and a cross-disciplinary approach to concept development. The show is in production for the Finnish Broadcasting Company.

4. ORGANIZERS

Marc Hassenzahl is Professor for "Ergonomics and User Experience in Design" at the Folkwang University in Essen. He is particularly interested in the positive affective and motivational aspects of interactive technologies – in short: User Experience. He is founding and active board member of the German Usability Professionals' Association.

Susanne Sperring holds an MA in sociology, BA in humanistic sciences, and is a doctoral student at Åbo Akademi University. She is the Manager at iDTV Lab, MediaCity, Åbo Akademi University, Vaasa, Finland.

Annika Wiklund-Engblom is a developmental psychologist, IT-pedagogue, and a doctoral student in the Research School for Media (Åbo Akademi University). She is the Research Coordinator at iDTV Lab, MediaCity, Åbo Akademi University, Vaasa, Finland.

5. REFERENCES

- [1] Hassenzahl, M., Law, E., & Hvannberg, E. 2006. User Experience - Towards a unified view [editorial]. In E. Law, E. Hvannberg, & M. Hassenzahl (Eds.), Proceedings of the workshop "User Experience – Towards a unified view" in conjunction with NordiCHI' 06, October, Oslo. Available online <http://www.cost294.org/>.
- [2] Hassenzahl, M. & Tractinsky, N. 2006. User Experience - a research agenda [Editorial]. Behavior & Information Technology, 25, 91-97.
- [3] Roibás, A. C., Geerts, D., Furtado, E., & Calvi, L. 2006. Investigating new user experience challenges in iTV: mobility & sociability. In Proceedings of the CHI 2006 Conference on Human Factors in Computing. Extended abstracts. ACM Press, New York, 1659-1662.
- [4] Tychsen, A., Newman, K., Brolund, T., & Hitchens, M. (2008). Cross-format analysis of the gaming experience in multi-player role playing games. In Situated Play - Proceedings of the DiGRA 2007, Tokyo. University of Tokyo, 49-47.

Workshop 3:

Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play

Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play

Marian F Ursu

NIM, Department of Computing
Goldsmiths, University of London
London, SE14 6NW, UK
m.ursu@gold.ac.uk

Pablo Cesar

CWI
P.O. Box 94079, NL-1090 GB
Amsterdam, The Netherlands
pablo.cesar@cwi

Doug Williams

Broadband Applications Research
BT
Ipswich, IP5 3RE, UK
doug.williams@bt.com

ABSTRACT

This paper provides a brief overview of the workshop on *Enhancing Social Communication and Belonging by Integrating TV Narrativity and Game-Play*. Its main objective is to explore new forms of TV-mediated communication between groups of people (such as family and friends) separated by space and/or time that would lead to improving the maintenance and fostering of their social relationships and, ultimately, of their feeling of being together. The investigation considers three perspectives: *socio-cognitive* – defining facets of the experience(s) of togetherness, when group interaction is mediated by the TV screen; *system design* – specifying requirements for TV-centric systems that support social interaction; *system implementation* – analysing enabling technologies.

Categories and Subject Descriptors

C.2.4 [Computer-Communication Networks]: Distributed Systems; H.1.2 [Models and Principles]: User/Machine Systems—*Human factors*; H4.3 [Information Systems Applications]: Communications Applications; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems; H5.2 [User interfaces]: Theory and methods; H.5.3 [Group and Organization Interfaces]: Theory and models; I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods; I.7.2. [Document Preparation]: Multi/mixed media; J.5 [Arts and Humanities]

General Terms

Design, Documentation, Experimentation, Human Factors, Languages, Theory

Keywords

Interactive Television, Social Television, Mediated Social Interaction, Mediated Communication, Narrativity, Game Play, User Experience, Presence, Personalisation, Enhanced Reality, Networked Media, Video Conferencing, Cinematography, Virtual Directing, Streaming, Multimedia Composition, Knowledge Representation

1. INTRODUCTION

Much research into *social interactive television* looked at the development of interactive services, offered whilst watching the TV content, that enable people to enjoy the social experience of TV even when they are in separate locations. Approaches include being able to share the viewed content – by seeing what friends are watching when programmes are broadcast or synchronising the watching of time-shifted content, sending recommendations and sharing favourites, or even sharing the remote control – and to comment on what is seen – by text chat, video conference or through avatars. Examples of prototype systems include AmigoTV (Alcatel-Lucent), CollaboraTV (AT&T), ConnectTV (TNO), and CoSe (Siemens); research endeavours along these lines have been presented and discussed, for example, at the Social Interactive Television Workshop at EuroITV 2007 [1].

In this approach, the TV programme, created by professionals, is the centre-point, and the other activities revolve around and are complementary to watching it. The programmes' performers are different from the audience. The TV programmes (both recordings and live transmissions) are distributed (e.g. broadcast) in a star model – from the distributor to each member of the audience – whilst the aforementioned complementary services provide for direct connections between the members of the audience.

What if the focus of the social interaction is shifted on other activities (i.e. not on watching TV programmes), such as social game-play, recounting of memories or even informal chats, between small groups of people who know each other (such as family and friends) but happen to be in separate locations? What if the professionally prepared TV content is replaced by an audio-visual communication, virtually directed, between them? Could the central role of the television screen be exploited when the performers become the same with the audience? Could the television's well established narrative forms and cinematic techniques be exploited in such a context? Could such an approach lead to a better sense of social belonging and togetherness? Could systems that support such forms of social interaction be built?

This workshop seeks to address these questions, and others that arise when exploring possibilities for devising enjoyable social experiences for people in separate locations when the screen is seen “merely” as a generic display device and not a “television” in its traditional sense.

2. AIM

This workshop aims to investigate how existing forms of social communication can be supported by, and new forms can be developed through, a *TV-centred communication*, with a view to enhancing the feeling of social belonging and togetherness between groups of people separated by space and time. More specifically, it aims to explore how family and friends, who are in separate locations, can share moments of fun together whilst playing social games, seeing and hearing each other whilst they laugh with and at each other, sharing instant impressions or recounting past memories, via a virtually directed TV-centred communication.

The workshop will have a particular bias on interaction and communication that is stimulated by or framed within *social game-play*. This should not be understood as merely attempting to transfer computer games onto the TV platform, by substituting the PC screen with that of the TV's, but rather as considering *the complete experience* of playing a game together: the game is just the pretext for socializing together; it is the reactions, the jokes, the laughter, the arguments, the parallel conversations, the recounted memories, and so on, that are in focus.

Social game-play could range from being very structured and rich in audio-visual content, such as electronic games, supported by a game engine, through structured but simpler in terms of interfaces, such as board games, to less structured and requiring no interface, such as guessing games.

The term *TV-centred communication* suggests that the audio-visual communication between groups of people goes beyond the standard face-to-face model of the current video-conferencing systems, aspiring to reach the aesthetic quality of good TV narratives, through employing cinematic techniques in both the capturing and editing of the content. It subsumes both *live* (real-time) and *catching-up* (off-line) communication. For live communications the content processing delays should not influence the fluency of the interaction.

In terms of the live communication, the aim is to create a *seamless connection* between the participating groups/locations. Each location, normally accommodating a *group* of people, must get the best account of what happens in (all) the other locations. This can be achieved only if there is significant automation in terms of both content capture and editing. This function can be denoted as automatic or virtual directing, or *interaction orchestration*. Cues, such as laughter, movement, talking person, ought to be extracted automatically, possibly helped by sensory information and direct instructions from the participants. They determine, on the basis of some embedded interaction or communication intelligence – denoted here as *orchestration intelligence* – the cinematic techniques that are to be applied for both content capture (cameras and microphones) and content delivery (screens and speakers).

Real-time orchestration can have two functions: to simply, but accurately and effectively, support the *recounting* of events happening in other locations; and to *moderate* the interaction.

The off-line communication is about semi-automatically constructing *TV narratives* which best capture key experiences. The content may be captured semi-automatically, during a live interaction, as described above, or manually, during events that happen outside such interactions (for example, as video recording or still pictures). Once the content was captured, TV narratives, possibly interactive, will be assembled semi-automatically. Such narrations could then be incorporated in live communications.

Orchestration, in this context too, has the meaning of a virtual director and editor: it is about deciding which content to record and select, and subsequently, about how to edit it in meaningful (interactive) TV narratives. This route is founded in the ShapeShifting Media Technology [2].

Scope. The setting is that of a limited number of households, each containing a *group* of people who know each other and want to stay connected. The focus is on interaction via moving picture and sound that are automatically captured and automatically edited. As output devices, the central role of the TV screen is preserved, possibly accompanied by secondary screens, but other devices, such as surround sound systems and ambient devices, may also be included. For input, each location will have a number of cameras, arrays of microphones and possibly other types of input such as from game consoles and sensors. All the devices should be such that they can be integrated in a household environment.

3. INVESTIGATION STRANDS

The proposed investigation considers three perspectives.

Socio-Cognitive and Perceptual – *defining facets of the experience of social belonging and togetherness.* Which are the facets of the experience of being together between people sharing a physical space at the same time, and possibly being engaged in social games? Which aspects of the social communication foster that feeling? Which could be transferred and supported through TV-mediated communication? Which TV formats and cinematic techniques (visual: types of shots, camera movement, edits, effects, etc.; and aural: spatial placement of sound, voice synthesis, sound effects, etc.) could enhance the experience of being together in a video conference like experience, but between groups of people in more than two locations? Are there new ways of communication, not possible without the technological support, which could foster the feeling of togetherness? Can there new communication techniques be proposed for *near* real time communication, when the delays due to audio-visual content processing are too large to be unnoticed?

System Design – *specifying requirements for TV-centric systems that support social interaction.* What related systems or prototypes already exist and how are they received by the end users? Which communication platforms do they employ? Which of their features could be adopted for the current aim? Are there new requirements/features refined through simulation and user evaluation? What is the economical feasibility of such propositions? What kinds of input and output devices are required, to implement this kind of social communication, and, particularly, emphasising TV based devices, what should their spatial layout be (in the household)?

Enabling Technologies – *analysing the capabilities of existing core technologies that could be employed in the development of TV centric systems for social interaction and, at the same time, refining new requirements for them.* Are there existing representation schemes which could (partly) express the envisaged communication intelligence for interaction orchestration? What are the capabilities of their associated reasoning techniques? Are there examples of AI mechanisms that could understand social situations related to the ones described here? Are there models which can *predict* aspects of social interaction? Are there any relevant personalisation techniques? Which features can be extracted automatically from audio-visual streams, possibly helped by sensory input and/or game states, and with what efficiency (processing time) and accuracy (precision)? What

description notations (ontologies) are there available? What multimedia notations for adaptive content exist? What are the capabilities of the low-delay decoding, transmission and encoding algorithms for both audio and video? What are the performances of the multimedia composition and rendering algorithms? How could game engines be integrated in TV based communication? How could different communication platforms be integrated?

The questions listed above are indicative: they guide but not restrict the proceedings of the workshop.

ACKNOWLEDGEMENTS

The organization of this workshop was partly supported by the Integrated Project TA2, Together Anytime, Together Anywhere (website: <http://www.ta2-project.eu>). TA2 receives funding from the European Commission under the EU's Seventh Framework Programme, grant agreement number 214793. The authors

gratefully acknowledge the European Commission's financial support and the productive collaboration with the other TA2 consortium partners.

REFERENCES

- [1] Lugmayr, A., and Gelobiowski, P. 2007. Interactive TV: A Shared Experience. TICSP Adjunct Proceedings of EuroITV 2007, Amsterdam (refer to the section on Social Interactive Television, edited by Geerts, D., Harboe, G. and Massey, N.)
- [2] Ursu, M.F., Kegel, I., Williams, D., Thomas, M., Mayer, H., Zombori, V., Tuomola, M.L., Larsson, H., and Wyver, J. 2008. ShapeShifting TV: Interactive Screen Media Narratives, *Multimedia Systems* 14 (2), ACM/Springer, 115-132.

ANSWER: Documentation, Formal Conceptualisation and Annotation of New Media

Angelos Yannopoulos

National Technical University of Athens
Dept. of Electrical & Computer Eng.
Division of Computer Science,
Zografou 157 73, Athens, Greece
ang@telecom.ntua.gr

Yannis Christodoulou

National Technical University of Athens
Dept. of Electrical & Computer Eng.
Division of Computer Science,
Zografou 157 73, Athens, Greece
yannis@telecom.ntua.gr

Stefanos Koutsoutsos

National Technical University of Athens
Dept. of Electrical & Computer Eng.
Division of Computer Science,
Zografou 157 73, Athens, Greece
stefk@telecom.ntua.gr

Katia Savrami

University of Patras
Cultural and Social Studies
Dept. of Theatre Studies
Rio 261 10, Greece
ksavrami@ath.forthnet.gr

Theodora Varvarigou

National Technical University of Athens
Dept. of Electrical & Computer Eng.
Division of Computer Science,
Zografou 157 73, Athens, Greece
dora@telecom.ntua.gr

Vassilis Alexandrou

National Technical University of Athens
Dept. of Electrical & Computer Eng.
Division of Computer Science,
Zografou 157 73, Athens, Greece
vassilis@telecom.ntua.gr

ABSTRACT

Within the scope of ANSWER, an EC-supported research project, a new artistic notation system, called DirectorNotation, is being developed for describing media content, much like notes are used to document a music composition. Our current research focuses on developing “notation-enabled” offline authoring tools, but interactive applications are an anticipated extension. Here, we summarise the key aspects of DirectorNotation and the ANSWER project, from the point of view of a multimedia-technology audience interested in content-based manipulation of media.

Categories and Subject Descriptors

I.2.4 [Artificial Intelligence]: Knowledge Representation
Formalisms and Methods – *Representations (procedural and rule-based), Semantic networks*

J.5 [Arts and Humanities]: *Arts, fine and performing, Performing arts (e.g., dance, music)*

General Terms

Design, Experimentation, Human Factors, Languages

Keywords

Artistic Notation, Film Directing, Knowledge Representation, Ontologies, Multimedia Annotation, DirectorNotation, Cinematography

1. INTRODUCTION

While technological support for the artistic activity of creating media is powerful, useful and ever improving, the available tools remain technical, and their interfaces continue to be shaped by implementation technicalities rather than a deeper artistic affinity with creative perceptions, concepts, needs and processes.

The EC-supported ANSWER project¹ [1] is a new approach to the creative process of film production. It will create a notation system [2], DirectorNotation [3], for describing the creation of multimedia content. The notation is a logical symbolic structure allowing directors to document their creative intentions, in a way similar to music notation and dance notation already used by composers and choreographers.

2. DIRECTORNOTATION - ARTISTIC AND TECHNICAL VALUE

DirectorNotation is interesting from a technological point of view. It is a system that simultaneously allows an artistic and a technical formalisation of the domain of film directing. Its main purpose is to analyse and represent media content during its production phase, rather than capturing knowledge about already existing media. The artistic notation is used by the film director to formally record his ideas and conceptualisation regarding the creation of a film (as opposed to using natural language or some kind of technical formal language similar to an animation script) – but it is also fully computer-processable. This leads to various applications – such as automatic synthesis of animated storyboards (a focus of ANSWER) or the automatic cost-estimation of a production from the earliest phases of pre-production (a topic of future research). All applications are based on the fact that DirectorNotation is transformed into metadata that accurately describes film content: we retain a machine-processable annotation of what the director was doing in order to produce the film. The description in notation is *formal* – the language’s semantics are strictly defined – so the automatically derived technical annotation corresponds exactly to the user’s input.

3. ANNOTATION OF NEW MEDIA

DirectorNotation has the potential to be exploited by any application that requires media content to be annotated. However,

¹ The ANSWER project has received funding from the European Community’s 7th Framework Programme (FP7/2007-2013) under grant agreement no 216489 (www.answer-project.org)

it can not produce annotations for every “thematic” point-of-view from which content could be described, so it is could not single-handedly allow e.g. for general-purpose consumer-oriented content-based media search [4][10]. The notation will describe a scene in terms of camera and actor motions, lighting and set configurations, spatial relations on-screen and in-set, dynamic qualities of movements, and more – this can be imagined as detailed, accurate *instructions* that the Director gives to his crew. Simple “interpretative” annotations of content as e.g. “happy”, “suspenseful” are not within its scope. However, annotations derived from notation have a very important advantage: they have been derived from the content-creator’s own description of the content (and ANSWER is developing a tool to match notation to produced footage so that the final annotation describes the final content exactly). This process is guaranteed to result in annotations that are detailed, accurate, and very extensive within the scope of the notation’s descriptive concerns – in clear contrast to the annotation of pre-existing footage, either manually or automatically by software.

DirectorNotation will be valuable for many new media and entertainment applications. For instance, the customisation of content that a system could perform when the content is delivered over an interactive medium will require a representation of how the content needs to be presented in different possible delivery scenarios. DirectorNotation allows both the artistic representation of a director’s ideas, and automatic processing of these ideas, so it can provide many new solutions in this area. For instance, [5] is based on “adaptive media templates” which allow amateur-produced content to be automatically edited according to predetermined professional designs, but it does not provide an artistically articulate way to define these templates, a gap that could readily be filled with our notation (and the corresponding ontology). We should stress that DirectorNotation will provide such applications with (formalised, well-defined) annotation *only*, not with processing algorithms. However, *combined* with content recognition algorithms, DirectorNotation would allow e.g. the “adaptive media templates” above to be applied to real-time interactive applications such as TV mediated communication.

4. A FILM DIRECTING ONTOLOGY

Our vision for the generalisation of the ANSWER project’s outputs is the creation of an “upper ontology for film directing” ([6]-[10]) that will offer developers a general-purpose, actionable knowledge model of the specific domain described above. When describing the architectural structure of content, based on decisions for the states and actions of camera, actors, lights and set elements, and decisions about film editing, the necessary vocabulary and semantics will be formalised in the generalised ANSWER knowledge model.

This model is intended to be put to uses that go beyond the scope of the ANSWER project. For instance, in our project, 3D graphics output is automatically generated, but this is always for the purpose of pre-visualisation, and thus not at “production quality”; also, the director always exactly describes his intentions, and the system never makes any decisions for the director. A future application, however, could aim to generate production quality graphics, or to provide automated decision making for the production of content, by adding the necessary modelling and reasoning components for issues such as narrative, aesthetics,

rendering, etc, over ANSWER’s formalisation of the domain of the film directing activity itself.

5. CONCLUSION

There is no space here to go into the technical details of how our vision will be achieved. The project website [1] is the entry point for additional information. A working software demo is intended to be released on the site by summer 2009. We should, however, note some important points in conclusion: all software developed by the project will be released as open source (and some legacy systems being brought on board all use open interfaces and open formats, e.g. a graphics engine based on X3D); our investigation of the director’s need for efficient description, as well as our application of film directing notation to creating video content in games, are leading towards the formalisation of idioms and rules, constructs which allow very efficient content description, and will be of special utility to further “reasoning” in applications that automate more of the content creation process; and, finally, we should close by crediting the system that inspired our project – Labanotation, the choreography notation [11][12].

6. REFERENCES

- [1] <http://www.answer-project.org>, valid Mar 2009
- [2] S.H.Waters, A.S.Gibbons, *Design Languages, Notation Systems, and Instructional Technology: A Case Study*, Springer Boston, Volume 52, Number 2, June, 2004
- [3] A.Yannopoulos, K.Savrami, T.Varvarigou, *Artistic Notation Systems Integrated in Software Engineering of Knowledge Technologies: emphasising film, animation and games*, ACM JOCCH (Journal of Computers and Cultural Heritage), under review / accepted with revisions, 2008
- [4] S.Bloehdorn et.al., *Semantic Annotation of Images and Videos for Multimedia Analysis*, Proc. 2nd European Semantic Web Conf., ESWC 005, Heraklion, Greece, 5/2005
- [5] M.Davis. *Editing Out Video Editing*. IEEE MultiMedia, 10 (2). 2-12.
- [6] B.Chandrasekaran, T.R.Johnson, V.R.Benjamins, *Ontologies: What are they? Why do we need them*, IEEE Intelligent Systems and Their Applications, 1999
- [7] N.F.Noy, D.L.McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, <http://www.ksl.stanford.edu/people/dlm/papers/ontology10/ontology101-noy-mcguinness.html>, valid Mar 2009
- [8] M.Uschold, M.Gruninger, *Ontologies: Principles, Methods and Applications*, Knowledge Engineering Review, Volume 11 Number 2, June 1996
- [9] T.R.A.Gruher, *Translation Approach to Portable Ontology Specification*, Knowledge Acquisition, vol. 5, pp. 199-220, 1993.
- [10] S.F.Chang, T.Sikora, A.Puri, *Overview of the MPEG-7 Standard*, IEEE TCSV, vol.11, no.6, pp 1-18, 1999
- [11] <http://user.uni-frankfurt.de/~griesbec/LABANE.HTML>, valid Mar 2009 (“Introduction to Labanotation”)
- [12] L.Wilke, T.W.Calvert, R.Ryman, I.Fox, *From dance notation to human animation: The LabanDancer project*, Journal of Visualization and Computer Animation 16(3-4): 201-211 (2005)

A Collective Director for Highly Interactive Viewing of Live Sports Events

Zhenchen Wang

School of Electronic Engineering and
Computer Science

Queen Mary University of London
London, UK

+44 207882 7408

zhenchen.wang@elec.qmul.ac.uk

Stefan Poslad

School of Electronic Engineering and
Computer Science

Queen Mary University of London
London, UK

+44 207882 3754

Stefan.poslad@elec.qmul.ac.uk

Alan Pearmain

School of Electronic Engineering and
Computer Science

Queen Mary University of London
London, UK

+44 207882 5342

Alan.pearmain@elec.qmul.ac.uk

ABSTRACT

New forms of personalised and interactive viewing experiences are being developed that enable spectators to act as directors of their own customised live sports videos. Scalability concerns for such systems dictate that group-based streaming rather than individual streaming can be realized in practice. In this positional paper, the application of the group personalisation model to enable social groups of possibly distributed individuals to collectively direct and share a common viewing experience of a live sports event is proposed. The design of this platform, developed as part of the My-e-Director 2012 project, is described and a discussion of a prototype system is given.

Categories and Subject Descriptors

H. Information Systems; H.5 Information Interfaces and Presentation (I.7); H.5.3 Group and Organization Interfaces; H.5.3 Group and Organization Interfaces.

General Terms

Design, Experimentation, Human Factors, Theory,

Keywords

Personalised, interactive, video streaming, group interfaces.

1. INTRODUCTION

Sports events are becoming increasingly instrumented, both actively by attaching tags to objects of interest, e.g., athletes, balls etc, and passively via vision based systems, enabling these objects of interest to be detected and tracked. Sports events can also be recorded from multiple perspectives using multiple cameras. Together these provide professional directors with a richer variety of content from which they can then select, schedule and broadcast a set of interlined views that they deem appealing to viewers.

However, different viewers may have different preferences and perspectives of what they find appealing in contrast to what is presented by the broadcast director. For example, different viewers may only be fans of and prefer only to track particular athletes based upon nationality or based upon recent performance, supporting, for example, only in form athletes. Multiple viewer preferences cannot all be accommodated by the director's cut of the event. Currently, sports events are not able to be directed and broadcast to match different preferences of different users.

The My-e-Director Project¹ [Poslad et al, 2009] is creating a new video delivery platform to offer viewers much greater choice and freedom to interact with live sports video content delivery. This platform improves social communication and belonging by putting the spectator in the director's seat. This enables spectators to choose amongst a wide range of camera streams, tracked objects and athletes, to use personal preferences, and for content to be delivered over a range of network infrastructures and access devices. Multiple views of the sports events with different tracked objects selected can be multicast over IP based networks.

The outline of this paper is as follows. First, related work is discussed. Next, the design and implementation of the platform is presented. Finally, conclusions and further work are presented.

2. RELATED WORK

Many projects have researched and developed personalised and interactive broadcast multimedia content delivery and access. ELIN focuses on personalisation of newspaper content. ePerSpace [Danet et al., 2006] focuses on building unified virtual personal environments and mCDN [Sokol and Eckert, 2007] focuses on personalisation to enhance the content retrieval process. IST-LIVE [Williams and Wages, 2008] also supports the concept of a future video conductor and user participation in real-time by allowing users to feedback to a director their choices for which parallel Olympic sports event they wish to view, enabling these to form 'non-linear stories' of events. Although IST-LIVE uses a similar concept to My-E-Director 2012, that of a director, the director concepts is used very differently. IST-Live seeks to support manual viewer feedback to schedule future events whereas My-e-Director 2012 focuses more on automated personalised viewing with respect to tracked objects of interest, on multi-camera use during live sports events that match personal profiles, and on the use of multicasts in addition to broadcasts.

3. MODEL OVERVIEW

The My-e-Director 2012 system consists of the following main components: video acquisition, video processing, camera selection, security, metadata generation and streaming, personalisation and terminal applications [Poslad et al. 2009]. Here the focus is on the personalisation component of the system. This acquires the individual personal preferences and clusters these to form group profiles.

The personalisation component is located between the video (and their associated metadata) stream server and terminal applications

¹ See also <http://www.myedirector2012.eu/>

within the overall system. It is used to direct the personalised video streams to multi-platform user terminals. This is achieved by clustering social group preference for sports event incidents (things of interest during sports events).

In My-e-Director 2012, sports event streams can be either manually or automatically switched. Streams can be predefined with respect to content focus (see Section 3.1), e.g. camera views of the pit in a long jump event defined as a stream focusing on a result change phase. Streams can be captured by a single camera or a set of cameras collectively. Spontaneous switches among these streams by spectators can generate personalised streams. The group personalisation processes forms two stages, i.e. **learning** and **inferencing**. *Learning* occurs after each viewing session if stream manual switch mode is enabled and *inferencing* occurs before each viewing session if user preferences for event incident types are given and the stream auto switch mode is enabled.

In inference stage, existing knowledge infers default rules for this user's preferences for event incidents and the system directs the associated user group preferred streams to that user. In the learning stage, a learning process updates the existing knowledge of user clustering. Figure 1 shows the workflow of the method.

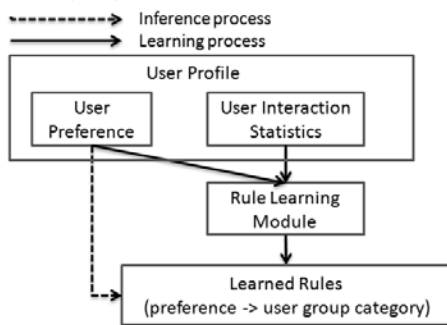


Figure 1 Social group preference clustering workflow

3.1 ACQUISITION OF USER PROFILES

Individual personal profiles are created and maintained via a mix of both explicit user input and through observations of user interaction.

User can explicitly express the preferences for a given set of predefined incident types for one particular sports event. The incident types are further grouped into three groups, namely **non-competition phase**, **competition phase** and **result change phase**. The defined phases are used to describe the content focus for the video stream and provide consequence vocabularies of clustering rules. Table 1 shows the defined incident types in corresponding groups. In practice, this allows users to select the interested incident types before viewing the sports event. Each incident type selected as assigned a value 1 or 0 if not selected.

Event	Non Competition phase	Competition phase	Result Change phase
Athletic Jump	Preparation, warm-up phases	Running, Take-off, Flying, Landing, On-your-mark	Distance announcement; Ranking announcement
100m	See above	Set, Running, Cross-finish-line,	Ranking announcement
Swimming	See above	On-your-mark,	One turn time

		Set, Jump, Swimming, Turn, Finish	announcement; Ranking time announcement
Football	See above	Kick off, Handball, Penalty kick, Offside, Red card, Yellow card	Score; Winner announcement

Table 1 Predefined incident types grouped in three phases

The system observes user interaction during viewing of a recorded sports event. Associated video streams are defined and associated with incident type groups. Users are allowed to switch to any interested recorded streams. The viewing duration for each defined phase is obtained with reference to the content focus of the stream and is recorded in an XML file (Table 2).

```

<Measure PName="Viewing Event" PDurationCounter="18" ClipTime =
"00:00:12.000" RecordingTime="28/03/2009 12:49:56" />
<Measure PName="Viewing Event" PDurationCounter="12" ClipTime =
"00:00:17.372" RecordingTime="28/03/2009 12:49:29" />
<Measure PName="Viewing Event" PDurationCounter="14" ClipTime =
"00:00:00.000" RecordingTime="28/03/2009 12:49:07" />
<Measure PName="Viewing Event" PDurationCounter="0" ClipTime =
"00:00:00.000" RecordingTime="28/03/2009 12:48:53" />

```

Table 2 Partial recorded viewing duration for a user

3.2 SOCIAL GROUP CLUSTERING

Social group clustering is done within the relationship learning module. A fuzzy neural network (FNN) framework is used to achieve this. Three social groups are defined, i.e. *non-competition process*, *competition process*, and *results prone*.

The motivation to use FNN is to address two critical issues concerning representing user preferences and group preference determination (i.e. rules determination). A neural network's leaning capability and a fuzzy system's fuzzy quantization power are both utilized. As a result, the constructed FNN for a social group clustering can be further decomposed into two major parts, namely a fuzzy membership function and the neural network.

People often use fuzzy labels such as "like", "not like" etc. to describe their preferences for one particular event incident. These labels can be quantized in a fuzzy system.

Another issue is that user interaction could reveal a preference that may not be directly reflected in explicit preferences. For example, two users may choose exactly the same incident type but the recorded viewing duration could be different. User A may be more result prone whereas user B may be more competition process prone. To solve this problem, different weights are assigned to each fuzzy label, e.g. 80% similar to a goal. A supervised learning neural network can do this by adjusting the expected weight of each fuzzy label given a training set, i.e. user profiles.

In order to produce the rules that cluster user social groups, the REFuNN [Kasabov, 1996] algorithm is used. Four modified steps are required to produce rules. They are user profile *fuzzification*, *FNN training*, *extracting initial set of weighted rules* and *aggregating the initial weighted rules*.

3.2.1 FUZZIFICATION

The user explicit preferences are fuzzified in this step, triangular membership functions ($0 \leq f(x) \leq 1$) are applied to determine the degree of interest on certain incident group. And three fuzzy sets (i.e. low, mediate and high) are defined for each fuzzy variable (i.e. non-competition interest degree, competition process

degree and result interest degree). The user chosen preferences within each incident category are summed and divided by the number of predefined incidents in that phase, i.e.

$$IGroup_i = \frac{\sum(Incident_{ticked})}{N} k, \text{ where } 1 \leq i \leq \text{group number}, 1 \leq N \leq \text{Incident number in phase } i \text{ and } k = \frac{N}{\text{Total Incidents No.}}$$

As a result, when the $IGroup_i$ is fed into the function, three membership degree are produced, e.g. if $IGroup_2 = 0.5$, where $0 \leq X \leq 1, 0 \leq Y \leq 1$, THEN membership degree for “low interest” would be 0, membership degree for “mediate interest” would be 1, and membership degree for “high interest” would be 0 (Figure 5).

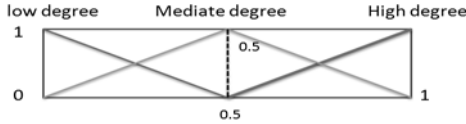


Figure 2 Fuzzification Example

Since user interaction data are expressed as crisp values already, they can be used directly as the expected results, associated with the explicit user preferences. A simple normalisation is performed by the following formula.

$$Duration_{max} = MAX \left(\frac{Duration_{group}}{Duration_{total}} k \right), k = \frac{Default Duration_{group}}{Default Duration_{total}}$$

3.2.2 FNN TRAINING

The previous step constructs the input vector and output result for the neural network. In this step, supervised learning is performed. The training set (i.e. user profiles) would be trained by a backpropagation training algorithm until convergence occurs.

The neural network is constructed with three layers. A number, e.g., nine, of input neurons are fully connected to the hidden layer neurons. The number of hidden layers (rule layers) is three which is the same number as output decisions (social groups). In this model, the sigmoid function with an alpha value of 1 is used to for each hidden neuron.

The following figure partially shows a trained network with 8 pseudo user profiles with a learning rate of 0.25, and trained for 1000 epochs. (Note that not all weights are labelled).

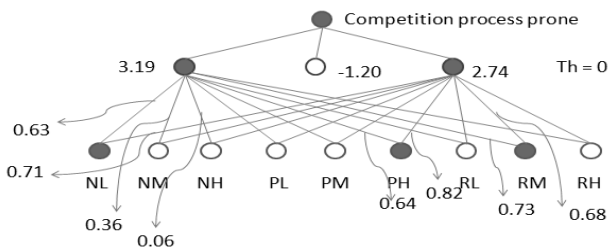


Figure 3: a partially trained neural network (ep =100, lr =0.25, n =8)

3.2.3 EXTRACTING & AGGREGATING RULES

Initial sets of rules are extracted from the trained network. Those connections from a hidden layer to the output neurons that significantly contribute to their activations are chosen. For each of these, its connections (those above a threshold, e.g., 0 in Figure 6) to a set of input neurons used in the antecedent part of expected rule. The connection to its output neuron is the initial value for the certainty degree. The partially trained network in Figure 6

demonstrates how rules can be extracted. Two initial rules obtained are:

R1: NL (0.63) & PH (0.64) & RM (0.73) \rightarrow Competition process prone (N, P, R stand for non-competition, competition and results; L, M, H stand for low, mediate and high interest degree)

R2: NL (0.71) & PH (0.82) & RM (0.68) \rightarrow Competition process prone

Note that the relationship for input variables is defined as synergism (collectively contributing to rule activation) in the network.

The last step is to aggregate the initial weighted rules with the same condition and consequent elements. The two rules obtained in step two can be aggregated by normalising the sum of initial degrees of importance of each condition element. i.e. $(R1+R2) / 2 \Rightarrow NL (0.67) \& PH (0.8) \& RM (0.71) \Rightarrow$ “More Interested in the Competition process”. Therefore, new user explicit preferences after fuzzification with fuzzy values can activate the rules that meet this rule.

4. DISCUSSION AND FUTURE WORK

This paper presents an approach to fulfil a collective director by grouping user preferences for event incident types. The proposed methodology can effectively learn the group preference clustering rules given the user preferences and user interaction data. The modelled neural work self explains the scalability of the system. Existing learned rules could effectively be updated once new set of user instances are added to the training set. Initial testing on the system shows that rules are produced as expected.

Future work could improve the current work by investigating issues such as multidimensional grouping, e.g. users can be grouped by other dimension such as athlete orientation, event type orientation etc. Issues concerning learning frequency and learning speed would also be investigated.

5. REFERENCES

- [1] Danet, P-Y., Madani, K., Kapovits, A. 2006. A Global Platform for Seamless Personalised Services at Home & Elsewhere - the ePerSpace approach. In Proc. 16th Biennial Conf. of Int. Telecommunications Society (Beijing, China 2006), ITS.
- [2] Kasabov, N.K., “Learning Fuzzy Rules and Approximate Reasoning in FuzzyNeural Networks and Hybrid Systems”, Fuzzy Sets and Systems, Special Issue, 1996:1-19
- [3] Poslad, S., Pnevmatikakis A, Nunes M et al. 2009. Directing Your Own Live and Interactive Sports Channel. Proc. 10th Int. Workshop on Image Analysis for Multimedia Interactive Services, (London, UK, 6-8 May, 2009) WIAIMS’09, Special Session on Event, Behaviour Video Analysis for Interactive Multimedia Services, in press.
- [4] Sokol, J., Eckert, K.-P. 2007. MCDN: Multimedia Content Discovery and Delivery. In Proc. 8th Int. Symp. on Autonomous Decentralized Systems (2007), 411-420.
- [5] Williams, M. Wages, R. 2008. Video Conducting the Olympic Games 2008: The iTV Field Trial of the EU-IST Project LIVE,” in Proc. 3rd ACM Int. Conf. Digital Interactive Media in Entertainment and Arts, (Athens, Greece, 10-12 September 2008) DIM.

(i)TV Brings People Together?

How the Feeling of Togetherness Can Be Build

Pauliina Tuomi

Tampere University of Technology
P.O Box 300
28101 Pori, Finland
+358503609803
pauliina.tuomi@tut.fi

ABSTRACT

Interactive television is becoming more and more socialized and it is no longer just a TV. There are numerous ways and technologies to interact with the TV (Parikka 2004; Jensen & Toscan 1999). There are also different viewpoints regarding what is interactive, what the level of interaction is, and furthermore, where today's television may be located on this dimension. What is iTV's role in building feelings of togetherness between people?

Categories and Subject Descriptors:

H.5.2 [INFORMATION INTERFACES AND

PRESENTATION]: User Interfaces – evaluation/methodology, user-centered design, graphical user interfaces (GUI)

General Terms: Design, Human Factors

Keywords: iTV-chats, cross media, togetherness, school shootings

1. INTRODUCTION

The current study of iTV entertainment (chats, TV mobile games, call quizzes) started in year 2004 and it consists of 30 hours of taped iTV formats and internet inquiry. (Tuomi 2008) The particular case-study (started in 2009) deals with the TV-chat broadcasted when the second school shooting occurred in Finland, 23.9.2008. The different aspects of togetherness were analyzed and discussed to answer these questions: What are the different most common motives and themes that create feeling of collectiveness? Which aspects of the social communication foster the feeling of togetherness? What is characteristic for this type of mediated interaction? What can be learned from it when developing better social communication systems? What different dimensions can be found in an abnormal situation such as media catastrophe? These questions were answered by analyzing the iTV field and text messages sent on the day of the school shooting. This analyzed data gives information how and why people might use TV screen as a collective and social meeting place.

Nowadays different technologies are becoming closer and closer to each other. Still, today's TV and the net are only converging at

some level. How is it possible that Finnish people chose chargeable TV-chat (approximately 1 €/message) over free crisis-chat on the net on the day another school shooting took place¹? Both, TV and the net, allow a lot of people to join in simultaneously and it is fairly easy to do so, using the keyboard or SMS-message. Both of them also offer the possibility to anonymity, the level to attend is not very high. It is risk free and it respects the privacy of the user. The biggest differences between the TV and the net are the use of the live host on TV-chats and the chargeability of these iTV formats, especially compared to chats on the net. (Tuomi 2008)

2. CONTEXT

2.1 Background

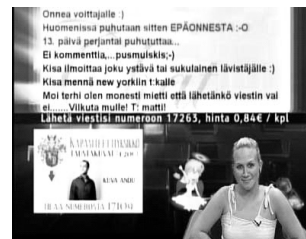


Figure 1. Human-Hosted TV-Chat

iTV-entertainment has quite a wide history. It has changed during the years but it is still going on strong which have probably surprised the game developers and TV-channels as well. At first, one could participate in different TV-chats (from the year 2000) – one could send greetings with a text message (SMS) and

almost immediately see his/her text on the TV-screen. See figure 1. This new form of entertainment soon became adopted by SMS game show producers. From the year 2002 there have been different kinds of TV-mobile games on television. At first they were games one could participate in with a text message – just by choosing the right coordinates to hit a certain target. Games were often based on problem solving and the interaction between the player and the game was limited to text messaging. Later on (2004), games developed further and a live human host stepped onto the playing field. For example, games were based on getting a football past the host or trying to hit her with a snowball. Since the participation took place via SMS or IVR², literally all viewers were able to participate. The level of interaction grew enormously, especially after the chat-function was added to the games. Now it

¹ There were crisis-chats on the net. For example: www.irc-galleria.net

² Interactive voice response is a phone technology that allows a computer to detect voice and touch tones using a normal phone call

was possible to play against a live host and talk to him/her – and most importantly: to get a response to one's action (Tuomi 2008).

iTV is still developing; most importantly via different convergences and divergences with other media technologies. With the advance of Internet and Web technologies, the increasing accessibility of mobile devices, the prevalence of rich media contents, and the ensuing social, economic, and cultural changes, technology and applications have evolved quickly over the past decade (Wang & Zeng et al. 79, 2007). Still, the basic iTV-formats can show us a way. The iTV experience, from social and individual point of view, is usually built on synchronous activity. Discussion forums on the net are based on asynchronous communication but this type of activity would not work on TV, because people especially want to attend and intervene on the broadcast flow. The possibility to attend and influence on a TV-format, especially real-time, is still, after ten years, something people seem to desire³ (Tuomi 2008). Instant interaction is best seen in cross media formats. Cross media entertainment means connection between mobile devices, Internet or/and television – the same contents are used in different platforms. Cross media enables for example gaming experiences between TV and mobile phones (Kangas 2004, 140). This paper focuses mainly on TV-chats and feeling of togetherness build on a TV-screen.

2.2 TV-Centred Communication?

What about the level of communication between the spectator and TV? How it has changed during the years and what could be predicted based on this knowledge? We are used to categorize communication and interaction at least from three different angles; interaction between humans, between a machine and a human and between machines (Isotalus 1998, 176). Interaction can traditionally be divided also into interpersonal communication, mediated communication and quasi-mediated communication. Usually TV and its stars and hosts are only trying to create simulation of interaction between spectators and the TV (Isotalus 1998, 177). Nowadays iTV-shows enable two-way traffic between audience and TV via chat-functions. (Tuomi 2008) All the previous research seems to pin point the fact that TV lacks in a real time-communication (Näränen 2006). That is not the case anymore, because people are able to communicate with the TV and a person on it.

That is why we should be talking about the fourth way of communicating which could be *real-time based mediated interaction/two way-synchronous communication*⁴ which highlights the interpersonal features of today's iTV. Of course we cannot define this form of communication interpersonal, not yet. Reasons for this are the facts that the TV-persons, in chat-function formats, are entitled to choose which messages they want to handle, the censorship of these formats limits the outcome seen from the TV-screen and you cannot force a TV-person to communicate with you by grasping his/her sleeve. The level of interaction has changed, but not yet to the point where we could talk about interpersonal communication. Anyway, the real-time based mediated interaction that chat-functions enable is a significant feature in today's interactive participation TV-formats.

³ Internet inquiry (35 questions and approximately 96 answers) concerning iTV-entertainment (2008).

⁴ Author's incomplete definition (Tuomi 2009).

3. THE ASPECTS THAT BUILD FEELING OF TOGETHERNESS

A qualitative study of creating feeling of togetherness and social interaction between Finnish people in this particular case consists of approximately 3 hours of taped TV-chat from 23. September 2008. The current research method is qualitative content analysis and the research is based on analysis of the numerous chat messages⁵. This research examines data as a case study; observation results and descriptive analyses make it possible to understand and define this phenomenon. This paper presents a classical case of collective behaviour stimulated by iTV. This is an interesting and worthy subject to study. Firstly, TV-chats could be seen a bit old-fashioned and they could have already vanished, which they have not. Secondly, TV chats being a place for sharing feelings of community and togetherness in 2008 is surprising. We have entered the era of web 2.0 which offers hundreds of social media applications where people can attend communities, almost for free (only costs are usually broadband-fees) The idea of social TV concept (including the old and newer inventions) will probably have to struggle with the net.

3.1 Chat-host as a Peer Support

Interactive entertainment has brought to life a new kind of TV-host culture. (Tuomi 2008) An iTV-host is very active and demanding on the viewers. The iTV-host is in the position of activating consumers to gain as many participants as possible. In this new iTV-host culture it is possible to communicate with the "TV-stars" by text messaging (SMS) and mainly because of this, there are different roles played by iTV-hosts. One role of the host is to be a peer support for the participating audience (Tuomi 2008). The dialog between the hosts and the players often consists of questions concerning dating and adult matters. Especially young people like to follow TV programs that are directed to people older than themselves and that handle issues of young grown-ups (Von Feilitzen, 2004).

"Mikko left me yesterday and I've been crying the whole day and night!" Host:" He wasn't worth it, tell him to jump off the cliff! Nelli (the host) is here as a relationship therapist" (MADX Tennis, summer 2006)

Still, it is also seen that older people rely on hosts' opinions when in doubt.

"I'm devastated! I caught my man cheating on me, again! What should I do???" (MTV3 Chat, 2007)

If participating costs could be reduced, iTV formats could turn into "legitimate" channels for advice and help (with the use of real experts). Such as public therapy, doctor's consultation, maternity clinic etc. There really is a need for this kind of a service, in health care for example⁶ (Tuomi 2008).

3.2 A Collective Space to Share Feelings

On the 23 of September 2008 Finland experienced yet another awful tragedy when the second school shooting took place in

⁵ Approximately 390 chat messages.

⁶ 18.07.2008: <http://www.poliklinikka.fi/?page=7874754&id=245669>

Kauhajoki. The killer was able to kill 10 innocent victims and himself at the end of the day. At the same time when the first news started to reach Finnish people, the *Mtv3 Chat* was on air. During this day and the next morning chats were full of messages concerning the shooting. Generally TV-chats became less of social activity since the host was added to formats. In the early chats people used to chat with each other and that is why chats were often described as a collective, social activity on TV screen. After the iTV formats started using human hosts, most of the messages started to be aimed at the hosts. (Tuomi 2008) All of this change when something similar to school shootings happens. So, this was not a new phenomenon, this had happened before in 2004 when the tsunami stroked and in 2007 when the first school shooting occurred. In 2008, people were still using TV chats as a collective place to grief and discuss. TV chats become places where feeling of togetherness is constantly reproduced.

SMS-messages from those days are easily categorized and this data shows us how collectiveness and togetherness can be produced on a TV screen and between people who do not know each other. SMS-messages from that day can be divided into: 1) feeling of horror expressed in the messages, 2) informative and questioning messages, 3) scapegoat and blaming messages, 4) solution messages, 5) aggressive messages, 6) messages from the people in Kauhajoki and 7) sympathetic messages.

1) Feeling of horror expressed in the messages

What has happened to Finland? Is killing really the only way to seek for attention?

2) Informative and questioning messages

Could someone please tell me, did anyone die there??

3) Scapegoat and blaming messages

TV and the net teach children violence. Or is Big brother suitable for children in your opinion?

4) Solution messages

Metal detectors and guards to schools, now!!

5) Aggressive messages

I hope the shooter will live.. God damnit.. We will put the boy pay for his actions!

6) Messages from the people in Kauhajoki

We, close relatives, don't know anything what is going on in that school where our family members are..

7) Sympathetic messages

Our deepest condolences! Family from Joensuu

TV chats have clear connection points to internet's discussion forums. They both have peer supportive role and value. All the messages above support community and mutual thinking patterns. People gather together as group on a TV screen against evil. Sense of togetherness can be build by sharing common grief. Sharing the grief and the shock often turn into symbolic action, which can be seen as media rituals (Couldry 2003). Ritual communication consists of producing, sharing and maintaining community (Carey 1994, 81-97).

4. CONCLUSIONS

This paper presented a descriptive study of iTV-sociality and collectivism among the audience. In conclusion, people get feeling of togetherness when they have a mutual agenda whether it is joy or grief. There are also ways to socialize in front of the TV with friends by playing game consoles. Today's console games seem to follow a certain pattern; social game-play. Nintendo Wii is based on social activity – let's play together. This can also be seen in different family/friends quizzes like in Buzz. This could be seen as a phenomenon when people are grieving after social communities, even in gaming situations. This could also be a new wave of arcade phenomenon where people gather together, even leaving their homes to play together and capture social gaming situations on a shared TV-screen. It is not rare bars and pubs to have Sing Stars et cetera. In conclusion, it is clear that people feel TV as some kind of a mystical machine, yet in 21st century. When attending or influencing on TV, it feels special. Whether it is sending SMS to the TV-screen or playing on it. We are, still, used to idea that TV is a one way broadcast medium and when being able to step on a flow; it feels as we are breaking a magic and public boundary (Tuomi 2008).

5. REFERENCES

- [1] Carey, J. 1994. Communication from Cultural Point of View. *Tiedotustutkimus* 17 (2): 81-97
- [2] Criteau, D. & Hoynes W. 2003. Media Society, Industries, Images and Audiences. Thousand Oaks: Pine Forge Press.
- [3] Von Feilitzen, C (Ed.). 2004. Young people, Soap Operas and Reality TV. International Clearinghouse on Children, Youth & the Media. Nordicom.
- [4] Herkman, J. 2001. Audiovisual Media Culture. Vastapaino. Tampere.
- [5] Isotalus, P. 1998. Television Performance as Interaction. *Nordicom Review* 19 (1): 175-183.
- [6] Jensen, J and Toscan, C. 1999. Interactive Television: TV of the Future or the Future of TV? Media & cultural studies 1. Aalborg University press. Denmark.
- [7] Kangas, S. 2004. Ubiquitous Computing in Game Applications. In Kankaanranta, M, Neittaanmäki, P & Häkkinen, P (edit.). *Digital Play Worlds*. Jyväskylä.
- [8] Näränen, P. 2006. Digital Television – Analyses on History, Politics and TV-system. Doctoral dissertation. University of Tampere. Tampere.
- [9] Parikka, J. 2004: The Interactivity's three Critics. *Lähikuva* 2-3, pp 83-97.
- [10] Tuomi, P. 2008. SMS-based Human-Hosted Interactive TV in Finland. *ACM International Conference Proceeding Series*; Vol. 291 archive (pp. 67-70). UxTV, Silicon Valley. California, USA
- [11] Wang F-Y, Carley K.M, Zeng D, & Mao W. 2007. "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 79-83.

Good to See You -

User responses to a TV-based video telephony service

Will Seager
Department of Computer Science
University College London
WC1E6BT
+44 207 679 3644
w.seager@cs.ucl.ac.uk

Hendrik Knoche
Department of Computer Science
University College London
WC1E6BT
+44 207 679 3642
h.knoche@cs.ucl.ac.uk

ABSTRACT

After many years of slow adoption within the market, video telephony is finally becoming accepted, e.g. in its high end version and in the form of video calls piggy-backed through computer-based VoIP services. Within the home, the TV set may serve as a useful additional way to access video-based communications. TV could also make video telephony accessible to people without computers or to people with insufficient computer skills. This paper presents a user study of a TV-based video telephony application. While, on average, more participants preferred the computer as a platform, many showed a strong preference for using video telephony on living-room TVs over computers, particularly those who experienced the application within a home setting. Our data indicate some strengths and weaknesses of living-room TVs as a platform for video telephony, and some requirements for the design of video telephony services aimed at the living room context.

Categories and Subject Descriptors

H.5.1 Multimedia Information services – Evaluation/methodology. H5.2 User Interfaces [user-centered design]

General Terms

Design, Experimentation, Human Factors.

Keywords

Video telephony, conferencing, TV

1. INTRODUCTION

With many commercial Voice Over Internet Protocol (VOIP) services now supporting video telephony (e.g. Skype), low cost video-based communications have become available to anyone with a computer, a webcam, broadband internet, and the computer skills required to install the software and connect the hardware. Increased availability has been followed by increased use. One quarter of the many millions of Skype-to-Skype calls each year now include video [1].

TV may serve as a useful additional or alternative way to access video-based communications. For people with computers, TV may provide a useful extra access point (e.g. when the computer is in use). In some instances, it may be the preferred access point. For example, users may prefer to make and receive video calls whilst seated comfortably in the relaxed and social atmosphere of the living room.

For people in households where there is no computer, the TV system may provide a low-cost platform for accessing services like video telephony. TV video telephony services may also help those with poor computer skills to access video telephony. Most people are already familiar with TV-based interaction styles; in particular, those with digital television are already familiar with using their remote to control interactive services. Provided that TV-based video telephony services utilize these existing interaction styles, those with little computer experience should find it relatively easy to use these services.

This position paper presents some findings from a user evaluation of prototype TV system that incorporated a video telephony service. During the study, user feedback was elicited in both laboratory and field settings. Our findings point to some of the strengths and weakness of the living TV set as a platform for video telephony services. Also indicated are some implications for the design of video services targeted for use within a living room context.

2. BACKGROUND

After AT&T's Picturephone failed to achieve critical mass in terms of market acceptance, subsequent research into audio-visual telecommunication focused on audio-video conferencing especially in work-related settings involving more than two parties e.g. [7], [4]. Products targeting professional settings include specialized video-conferencing systems such as HP's high-end telepresence system 'Halo'. For the most part, the problems addressed by this research do not necessarily apply to video calls in private settings. For example, the problem of representing more than two parties [2] visually might not be as important in private video communication settings, which are usually one-to-one. There is a growing recognition of the importance of understanding people's needs with respect to technology use in the home [6], [3]. In contrast to research investigating video-based communications within work settings, video telephony within the home has received little attention. Our study gathered user responses to a TV-based video telephony application designed for a home setting that was evaluated both in the lab and home contexts.

3. METHOD

The prototype TV system evaluated in the study was developed within the Universal Satellite Home Connection (UNIC¹) project. It was trialed during March, April and May 2008. Apart from the video telephony services, the system provided internet browsing, chat, television, video-on-demand, music, an electronic programme guide (EPG), news, radio podcasts, weather, and a service for sharing user-generated content. All services were delivered via a 2-way satellite-based infrastructure to a set-top-box and TV set. Users interacted with the system via a remote control and wireless keyboard and, when using the video telephony service, via a webcam. Laboratory trials were conducted in UK, France and Italy. 27 participants (avg. age 31) were asked in 90 minute individually run trials to fill in a questionnaire and carry out scenario-based tasks with each service. They were also interviewed about their general impressions of the system. Field trials were carried out in 6 households at sites in UK, France, Italy and Germany. Nine field-trial participants (avg. age 48) and their families were asked to use the system at their own convenience. The field trials varied from 10 days to 4 weeks. At the end of each trial, semi-structured interviews elicited general impressions of the system, responses to particular features and identified usability problems. Figure 1 shows the UNIC video telephony service in part-screen mode.



Figure 1: UNIC TV-based video telephony service

4. RESULTS

In the post trial interviews, participants were asked to imagine that they had a UNIC-like system with multiple TV-based interactive services and a computer with internet access, and were asked to indicate how they would like to access each service. Figure 2 shows the preferred platform for the video service.

A slightly higher proportion of laboratory participants preferred to access the video service via their computer (32% as opposed to the 28% who preferred the TV platform). However, the pattern amongst home trial participants was reversed with more people choosing TV as their preferred platform.

For the lab-based participants, it is notable that a large minority either preferred the TV platform (28%) or indicated they would use computer and TV equally to access video services (41%).

¹ <http://www.unic-project.org/>

It should also be noted that all the laboratory participants had a computer at home. Possibly, people without computers would express a stronger preference for the TV platform.

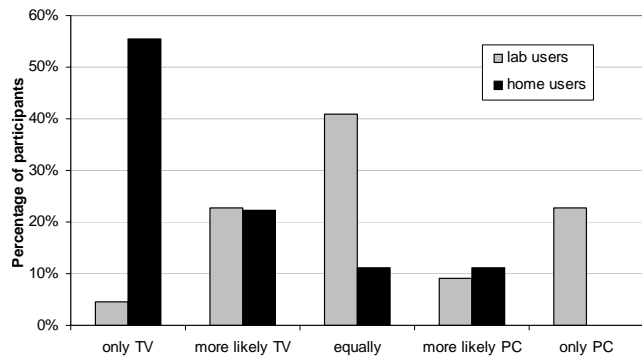


Figure 2: Preferred platform for audio-visual calls

4.1 Pros and Cons of TV-based audio-visual communication

While experiencing the UNIC video phone service, or in the post trial interviews, some participants pointed to advantages of the living room TV set as a platform for video telephony services. A frequently cited advantage relates to the fact that, for most people, the living room is a place to relax and socialize. *“When you have a phone conversation where do you normally talk? You normally sit in an armchair relaxing talking to your friends. So I would say the most logical place to have [a video phone service] would be in that sort of relaxing talking setting”.*

An advantage for some participants concerned the close proximity of the living room to other social areas of their household, in particular the kitchen. With the living room TV closer to areas where they spent most time during waking hours, it would be easier to make and receive video calls. Also, ease of access was helped by the fact that the TV was often switched on.

Several participants thought the TV particularly well-suited to conversations where more than one person wanted to take part at one or both ends. TV sets typically have larger displays, and the displays can usually be seen from several comfortable seating positions. *“With a big screen in the living room...the whole family can sit down and potentially interact...you can see everyone at the same time and they can see you...with a computer, it's more restrictive”.*

The main perceived drawback of video calls via the living room television was the potential for reduced privacy and freedom from distraction given the increased likelihood of others being present or entering the room during a call.

Some favoured the TV set for conversations that involved friends and family but not for business or work related conversations. For the latter, they thought it would be easier to avoid disruption when using the computer.

Some were also concerned about potential competition for use e.g. when they wanted to make or receive video calls, other household members might want to watch television or play games.

4.2 Requirements

One key requirement expressed by many participants was that the system should provide safeguards for their privacy. Many were

concerned that hackers could potentially gain access to the webcam and, thereby, view what was occurring within their living room. For some, this was a reason enough not to have a video telephony service in the living room. Others felt that their privacy concerns would be assuaged by a physical means of blocking the camera e.g. several suggested a plastic moveable cover that could be brought down to block the camera view – featured by a number of webcams but not the one used in the trial.

Many participants were keen to have the type of online/offline status indicators provided by existing VoIP services. In particular, many wanted an ‘invisible’ status indicator so they could view who was online without being visible. Some were keen to set which status indicator appeared when switching on the TV system and when entering the video telephony service.

Another common concern was that incoming calls could disrupt television viewing. Some wanted audio rather than visual alerts for incoming calls arguing that visual alerts would be too distracting. Also, audio warnings would be better for alerting people who were not directly watching the television.

Several said they would want the option to ignore incoming calls – e.g. when engrossed in film or television programme – and believed this would not be easy with the status indicators provided by computer-based VoIP services. Online indicators suggested you were at home and, thus, made it difficult to ignore calls. Offline or invisible indicators, on the other hand, suggested you were not contactable at all. One suggestion was to dispense with status indicators entirely: as with traditional phones, it would be easier to ignore calls because it would not be obvious that you were at home and ignoring the call.

Several participants said that it should be possible to make or receive calls when the TV system was switched off. One suggestion was to route incoming video calls via other phones as audio calls when the TV system was switched off.

Some participants noted that, to be usable with friends and family, the service would need to be interoperable with other video telephony services, as they did not think it likely that many friends and family would share the same TV system.

One participant pointed to a problem that often occurred with computer-based video telephony services: the camera of the person at the other end of the call could be zoomed in or out too far or not pointed in the right direction. He thought this was more likely to occur in the living room context as people would be further away, in differing seating positions and because, sometimes, there would be more than one person at the other location. He suggested a potential solution to the problem: each person in the call could have control of the camera at the other end. By moving the camera, they could zoom in on the other person’s face to see facial expressions or zoom out to see the wider context. They could also scan the camera if someone else was present.

5. DISCUSSION & CONCLUSION

In summary, our findings suggest that many people would like to access video telephony services via their living room television set. Perceived strengths of the living room TV platform include the relaxed and sociable atmosphere of the living room, and the fact that the living room TV set is usually viewable from multiple seating positions and thus better suited to shared use at one or

both ends of the conversation. Perceived drawbacks include a potential for reduced privacy and freedom from distraction, and the increased likelihood of competition for use of the living room TV display compared with a computer.

Our findings also indicate some requirements for video telephony services targeted at living room TV sets and, in some cases, potential design solutions. Many users were keen to have a means to prevent unauthorized access to the webcam. A physical barrier blocking the camera was viewed as a good way to achieve this. Another common requirement was to have mechanisms to help users avoid distraction from incoming calls. The main suggestion here was for audio alerts and for status indicators that invited calls but were easy to ignore if necessary. Possibly, ambient awareness solutions as implemented in [5] would help meet this requirement. Some users were keen to have interoperability with other devices in the home (e.g. telephone) and with other kinds of video telephony service. Finally, there may be a need for mechanisms to allow users to switch between detailed and context views e.g. to support conversations where there is more than one person in the living room at the other location. A solution might be to allow users to control the camera at the other location (see for example Polycom’s Video Conference Systems). Alternatively, automatic camera movement and/or switching, as suggested in [7], might provide a solution. Whether these approaches, in which people relinquish control over their camera, are compatible with the aforementioned need for privacy would need to be tested in people’s homes.

6. REFERENCES

- [1] Arak, V. *Interview with Skype CEO Josh Silverman*. Available at: http://share.skype.com/sites/en/2008/04/interview_with_skype_ceo_josh.html, 2009.
- [2] Buxton, W. A. S., Sellen, A. J., & Sheasby, M. C. *Interfaces for multipoint videoconferencing and its implications for design*. In A.J.Sellen, S. B. Wilbur, & K. E. Finn (Eds.), *Video Mediated Communication* (p.385-400). Lawrence Erlbaum Assoc Inc 1997
- [3] Edwards, W. K. & Grinter, R. E. *At Home with Ubiquitous Computing: Seven Challenges*, in *Proceedings of the 3rd international conference on ubiquitous computing*, p.256-272, London, UK: Springer, 2001
- [4] Fish, R. S., Kraut, R. E., Root, R. W., ice, R. E. *Video as a technology for informal communication* *Communications of the ACM*, 36 (1), p.48-61, January 1993.
- [5] Harboe, G., Metcalf, C. J., Bentley, F., Tullio, J., Massey, N., Romano, G. *Ambient social tv: drawing people into a shared experience*, in *Proceedings of CHI’08*, p.1-10, New York, NY, USA: ACM, 2008
- [6] O’Brien, J., Rodden, T., Rouncefield, M., Hughes, J. *At home with the technology: an ethnographic study of a set-top-box trial* *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6 (3), p.282-308, 1999.
- [7] Ranjan, A., Birnholtz, J. P., Balakrishnan, R. *Improving Meeting Capture by Applying Television Production Principles with Audio and Motion Detection* in *Proceedings of CHI’08*, p.227-236, New York, NY, USA: ACM, 2008

A Multimedia Content Semantics Extraction Framework for Enhanced Social Interaction

Giuseppe Passino, Tomas Piatrik, Ioannis Patras and Ebroul Izquierdo
Queen Mary, University of London
Mile End Road, London, E1 4NS, UK
<name>.<surname>@elec.qmul.ac.uk

ABSTRACT

In this paper, a system for improved social interaction via the Internet or interactive TV is proposed. The aim is to provide a small group of closely connected users with a rich social experience, sharing intimate moments of life and emotions, taking full advantage of the existent Internet technology and broadcasting practices. Starting from a single use-case, a feasibility study for a social interaction is illustrated. The proposed architecture for social interaction is based on techniques for automated extraction of semantics from streamed content. In particular, technical feasibility and real-time processing issues are considered. Semantic information is used in the multimedia editing and composition phase, enabling the system to offer an experience that goes beyond the classical face-to-face video-conference. The efficient and rich presentation of the content is driven by technology for semantic segmentation, object detection and automated extraction of interesting regions in the scene. Furthermore, a face detection module is used to guarantee a constant visual presence of the parties. Finally, a summary of the session is automatically generated for future uses or on-line-browsing during the conversation.

1. INTRODUCTION

The pervasiveness of Internet connectivity is generating a social revolution. Users are experiencing deeper interaction independently from remoteness and geographical locations, and it is expected that the use of such remote interactions will continue to grow in the years to come. In this paper we argue that it is possible to integrate classical content-delivery technologies as broadcasting, optimised over many decades, with innovative computer vision techniques to produce a socially rich communication environment through semantics-aware systems, *i.e.*, systems able to react according to the semantics of the streamed content. We present a proposal for an integrated system to enable a rich user interaction experience based on existent broadcasting technology and advanced semantic multimedia analysis.



Figure 1: Output video example for the considered use-case.

The scenario considered in this work entails rich social experience between few closely related people, in different locations. The communication is carried out through a multimedia channel, exchanging content acquired through multiple video inputs, audio and complementary user sources. The goal is an immersive experience that goes beyond the classical face-to-face videoconferencing. To illustrate the underlying idea in an effective way the following scenario is used: a young student living abroad is interacting with his mother at home regarding a cake recipe. Using one or more conventional video cameras, the mother can share the preparation of the cake from her own kitchen while communicating in a friendly and close atmosphere, and the content is offered to the son using advanced automated video editing and presentation techniques typical of television, such as zoom and framing effects. The face of the participant may be always visible in a corner of the display, being the face automatically detected and framed by the camera. Kitchen tools, and other interest objects in the scene will be isolated and focused via real-time object detection algorithms. Finally, an automatic summary of the entire video is produced for efficient and easy on-line fruition and later viewing. In Fig. 1 a possible output example is shown.

The remainder of the paper is structured as follows. In Section 2 we present an overview of the proposed system; in Section 3 the face detection problem is presented, while the general and diverse problem of interest object detection is discussed in Section 4. Scene detection is treated in Section 5, and automatic summarisation in Section 6. Expected performance of the components and conclusions are presented in Section 7.

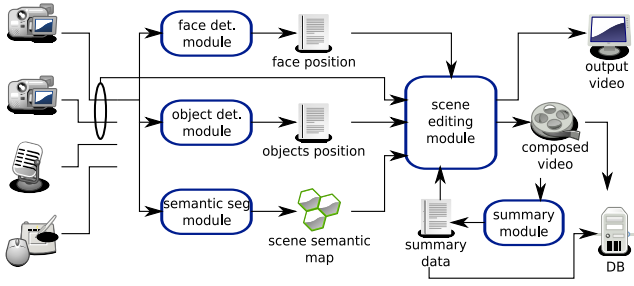


Figure 2: Block diagram of the framework as detailed in Section 2.

2. SYSTEM OVERVIEW

In this work we explicitly address the multimedia processing activities related to semantics extraction, while other aspects of the system are omitted for the sake of conciseness. We do not elaborate on the module for automated content editing and presentation, that should implement techniques borrowed from the TV production world. The basic system modules related to semantics extraction are a face detector, a generic object-of-interest detector, a scene semantic segmentation module, and an automatic summarisation module, as depicted in Fig. 2.

Eye-contact plays a crucial role in communication, for effectively conveying messages and sharing feelings: the face detector component is aimed at spotting the position of the face of the parties, enabling its zooming, framing and constant presentation in a complex final layout. On the other hand, in order for the editing system to react to the current activity, awareness of the objects present in the scene is needed. This requirement is satisfied with a generic object detection algorithm, independent on the particular nature of the object of interest. Finally, a more semantically aware component aimed at the understanding of the entire scene is used to link concepts related to the discussion and the content of the scene (for example, in a kitchen environment, the location of people, cupboards, oven and other appliances can be relevant to the discussion). The content is composed by an automatic content editing component which receives as an input the audio-video streams and the associated semantic information extracted by the detectors. Finally, a summarisation module, applied to the output of the editing block, will automatically produce a set of highlights to easily browse the content, either on-line during the conversation, or for a later usage (for example, to revise the steps of the cake preparation).

3. FACE DETECTION

In the analysed scenario the face detection is not particularly challenging, involving a restricted number of people in an environment with low dynamics. Different optimised algorithms have been used in end-market products, even for portable devices as digital cameras. An approach based on boosting [9] allows for real-time detection. To improve location performances a facial model [1] can be used. A probabilistic model based on a fixed number of stable points is created. The model is then considered as a flexible structure that is efficiently positioned in the best location in the vicinity of the location for the detected face obtained through

the boosting-based detector. The detected faces are then tracked with a standard method [6].

4. OBJECT DETECTION

The object detection module is aimed at the identification of the object of interest, in order to focus on it during the video editing stage. This is the most challenging aspect of the framework, since the nature of the objects is unknown by definition, and therefore the object detection module has to be generic: an explicit search for a large set of object categories is not viable in real-time, and its training demanding. For the previously stated reasons, we propose the use of a semantic latent method, in which the categories are not explicitly specified but are instead hidden. We base our approach on the success of the bag-of-words model to detect and localise objects in images [7]. This method is based on probabilistic Latent Semantic Analysis (pLSA), where (visual) words are associated to latent object categories. The categories are therefore described in terms of word histograms. The visual words are salient points in the image encoded in a dictionary of features. The “quality” of the words, as the robustness to image geometric transformation and descriptivity, is affected by the real-time constraint, but the extraction step is generally feasible within these limitations. The adaptive (unsupervised) training step can be done in background and updated at regular intervals.

The output of the pLSA is a set of probabilities of words given the latent categories. From this information we can estimate the category posteriors given the visual words. This is a weak information on the presence and location of an object, as local concentration of words related to a single topic indicate the presence of an object of interest. Since salient points are robust against geometrical transformations, a tracking strategy [6] can be applied to estimate the degree of movement of different objects in the scene.

5. SEMANTIC SCENE SEGMENTATION

The semantic segmentation module is demanded to give an interpretation of the layout of the scene, associating a semantic category label to each image pixel. We propose to use a method developed in our research group, in which the image is analysed by extracting parts and considering their aspect and context [4], with some modifications to respect the real-time constraint required by the application.

The basic module activity can be summarised as follows: image parts are extracted, and descriptors are evaluated for each part; an optimised structure is then computed, to connect the parts; finally, based on the calculated structure and the part features, the parts are labelled through the application of a Conditional Random Field (CRF), a probabilistic graphical model.

This this scenario, since accurate border detection for object is not necessary, a fast part extraction strategy based on a grid of rectangular patches is proposed. The descriptors, based on textons [2] and hue histograms [8], rely on visual dictionaries, that are expensive to compute but need to be compiled offline only once. The optimal structure evaluation is based on the Aspect-Coherent Minimum Spanning Tree (AC-MST), whose associated complexity is quasi-linear, being a MST algorithm. The discriminative nature of the CRF

Table 1: Illustrative pixel-level labelling performance for the semantic segmentation module.

Building	Grass	Tree	Cow	Sky	Airplane	Face	Car	Bicycle	Average
63.0%	94.2%	68.9%	84.4%	93.7%	75.8%	92.9%	76.4%	86.5%	82.9%

makes the inference straightforward in terms of time performance, but the model needs to be trained with pixel-level ground truth, in a time-consuming process. This can be done on-line when installing the cameras, with the help of the user to provide partial image annotation, which is supported by the method.

6. AUTOMATIC SUMMARISATION

The process of creating video summary includes three important steps, namely shot boundary detection, key-frame extraction and shot relevance/redundancy estimation. Many efficient techniques have been proposed to deal with the aforementioned tasks, however most of them are not dedicated for real-time applications because of their high complexity [10, 3]. To reduce the computational complexity as required in this application scenario we propose an approach which combines both scene detection and key-frame extraction to create the summary excluding redundant segments.

A fundamental step in our approach is to create the similarity matrix and organise video frames into a tree structure using Ant-Tree Strategy (ATS) [5]. ATS is inspired by self-assembling behaviour of ants and their ability to build mechanical structures. On the basis of a root on which the tree is built, frames are gradually fixed to the structure. The movement and fixing of a frame in a specific position depends on its visual features, temporal information and the local neighbourhood of moving frames. Results of the ant-tree algorithm are clusters used in the decision process for classification of relevant/redundant segments. Common video segments which contain representative frames attached to the root of the tree are classified as relevant. The importance of relevant segments is defined by the number of frames from redundant segments in the corresponding cluster. Finally, a video summary is constructed by concatenating relevant video segments or representative key-frames. Additionally, the obtained summary is progressive, with richer descriptions deeper in the summary tree.

7. RESULTS AND CONCLUSIONS

In the early stage of the proposal, not having a complete system prototype yet, we could not perform any comprehensive test. However, general results for the modules allow us to draw preliminary conclusions with a good level of accuracy. For the scene detection module, time performance is not an issue, since it has already been shown how the algorithm works in real-time: without the help from the tracking component, the face detection algorithm runs at 15 frames per second on a modest consumer PC [9], with high associated detection performance (around 92%-94%). The performance of the generic object detection system is difficult to evaluate, since in this scenario this task is rather subjective. Published results related to the proposed method validate the approach by classifying images containing only one out of four general categories (car, aeroplane, faces or motorbike), or background [7]. The results show an average classification rate of over 93% in this relatively simple task. The

more challenging object localisation task, evaluated for the face category, achieves a poorer 60% accuracy, still giving useful hints on the relative position in the image.

Semantic segmentation has been tested the Microsoft Research Cambridge (MSRC) dataset of nine categories. The system can however be easily tailored on the specific scenario to be analysed, either indoor or outdoor. Indicative performance obtained with the base model on the MSRC dataset are presented in Table 1. Finally, we have tested our summarisation algorithm as a part of the collaborative system for automatic video summarisation in the TRECVID 2008 BBC rushes evaluation. Evaluation criteria included: *IN* – the fraction of inclusions found in the summary ($0 \div 1$); *JU* – “the summary contained lots of junk” (1 strongly agree – 5 strongly disagree); *RE* – “the summary contained lots of duplicates” (1 strongly agree – 5 strongly disagree). Our method scored: *IN* = 0.4, *JU* = 3.4 and *RE* = 4. The results show how the semantics that can be extracted from the video streams achieves an accuracy that enables an effective use in video editing scenarios. The system is highly expandable and audio processing data can be integrated, for example, for a richer scene semantic analysis.

8. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Seventh Framework Programme under grant agreement no. FP7-216444, PetaMedia.

9. REFERENCES

- [1] M. Everingham, J. Sivic, and A. Zisserman. Hello! my name is... buffy – automatic naming of characters in tv video. In *BMVC*, 2006.
- [2] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.
- [3] A. G. Money and H. Agius. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation*, 19(2):121 – 143, 2008.
- [4] G. Passino, I. Patras, and E. Izquierdo. Image semantic labelling via heterogeneous low-level descriptors integration in a conditional random field. In *WIAMIS*, 2009.
- [5] T. Piatrik and E. Izquierdo. Hierarchical summarisation of video using ant-tree strategy. In *CBMI*, 2009.
- [6] J. Shi and C. Tomasi. Good features to track. In *CVPR*, 1994.
- [7] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering objects and their location in images. In *ICCV*, 2005.
- [8] J. van de Weijer and C. Schmid. Coloring local feature extraction. In *ECCV*, 2006.
- [9] D. Zhang, S. Z. Li, and D. Gatica-Perez. Real-time face detection using boosting in hierarchical feature spaces. In *ICPR*, 2004.
- [10] S. Zhu and Y. Liu. Automatic scene detection for advanced story retrieval. *Expert Systems with Applications*, 36(3, Part 2):5976 – 5986, 2009.

Experience, narrative and interaction in TV-centred communication

Frank Nack

HCS – University of Amsterdam
Science park 107,
1098 XG Amsterdam, The Netherlands
Tel: +31 20 525 6377
E-mail: lastname@uva.nl

ABSTRACT

Based on the view that changes in user expectations and social demographics alter the traditional use of TV this paper investigates the notion of experience, narrative and interaction in TV-centred communication. We have a look at synchronous and asynchronous consumption of expert as well as user-generated content. We then investigate participatory elements that enhance the feeling of social belonging and togetherness in particular when TV users are separated by time and space.

Categories and Subject Descriptors

H5.4 [Hypertext,Hypermedia]: *Architectures, Navigation, User issues*. I.7.2. [Document Preparation]: *Hypertext/hypermedia, Multi/mixed media*.

General Terms

Human Factors.

Keywords

TV-centered communication, experience, narratiity, TV genres, interaction patterns,

1. INTRODUCTION

The TV screen is nothing more than a window to access information, which, not necessarily but often, is presented in narrative form. Due to changes in the public's viewing behaviour, mainly fostering choice and interaction, as well as changing social demographics, i.e. temporal or spatial separation of people in close social relationships, there is currently much talk about expanding TV's typical one direction information stream. The aim is to facilitate a more participatory way of TV consumption, including and in fact

focusing on well-known forms of social interaction, to enhance the feeling of social belonging and togetherness in particular with those who do not share either time or space with the gatherers around the TV screen. Thus, the aim is to turn a common communication receiver, particularly used as a source of entertainment and news either in form of recorded or broadcasted material, into an interactive medium that re-establishes the lost social world. In this paper we investigate the various options and their technological requirements to support richer social interaction in a TV-centred environment. We emphasise in particular on participatory elements inherent in time, space and their influence on the representation and presentation of the content.

2. COMMUNICATIVE TV - DEFINITIONS

The technological environment to be considered in this paper consists of large flat screens as they can be found nowadays in a great number of private homes and a connection to the outside world, either through cable, setup-up box or satellite. The aim of this environment is to provide content that is generated by experts for the distribution through broadcasting, or by non-professionals where the distribution channel will be most likely web oriented. In both cases the goal is to entertain or inform.

Professionally generated content, often mirrored with respect to structure or aesthetics by non-professional content, follows in most cases distinct narrative notions [1]:

- A narrative is restricted to a more or less stringent chronological model of temporal experiences (e.g. suspense and expectation), where beginning and end have been determined and such unfolding a destiny rather than a passage into uncertain territory. Typical genres here are drama or comedy covering crime, horror, or thriller, western, science fiction, or situation comedy (sitcom). Due to their complex structures these genres usually require some authoring expertise, which makes it complicated for amateurs to establish them. However, users are willing to engage with provided

narratives, i.e. by navigating through the story space if they are assisted [2, 3].

- Almost any mode of human behaviour can be understood as narrative, representing essentially narrative forms of memory, myth, history, news, politics, or science. Those are reflected in TV genres such as news, documentaries, daily soap operas, talk shows and reality TV, and so forth. As these formats require less structured continuity there are several mechanisms how computer can automatically generate stories in such genres based on user generated content [4, 5].
- A narrative as a poststructuralist inspired attack on the dominant culture (see the two bullet points above) in form of oppositional, resistant or subversive material. Typical genres here are avant-garde documentaries or comedy shows. These types of program can be also be established in an automatic fashion, based on user generated content [6, 7].

Whatever the content, the way in which it is consumed is passive, semi-active or interactive. Viewers can sit alone or gather in groups around the screen and consume the material at the same time – or consume the same content in either a spatial and/or temporal asynchronous way. In fact, narrative universes can become reversible and no longer need to reflect the psychological cause and effect as freed social participatory environments facilitate people to consume different programs a- or synchronically and communicate exactly about that.

The essential aspect of the narrative, seen as an event, is that it establishes an experience in the audience. In the context of TV-centred media distribution we understand an experience as the nonidempotent alteration of the cognitive map and/or related cognitive processes of the one who has the experience, derived from direct observation or participation in a contextualised event or activity over a certain period (short or long-term) [8].

It is in particular the different ways of social interactions, in different temporal and spatial situations we will focus on in the ongoing discussion.

3. INTERACTION MODES IN COMMUNICATIVE TV

The essential elements that have a large influence on the users expectations with respect to expressiveness of the content as well as their willingness to actively engage or passively consume depend mainly on the idiosyncratic personality of the individual viewer. The temporal and spatial context will, however, influence his or her behaviour substantially. In this section we investigate the synchronic and asynchronous relationships between the concepts time, space and content and their inherent

possibilities and challenges for communicative and experiential TV. Their role for user modelling is mainly neglected in the discussion, even though this point is relevant. We also do not address the classical notion of watching TV at the same time, same space with the same content, as this is the state of the art.

3.1 Same time, same space, different content

This is the classical situation of people gathering in the living room for watching TV, only that here we address the current development of multiple screens in one location on which different programs can be enjoyed at the same time. As people share the same experiential space they will make use of natural observation techniques to share their experiences: they look at or talk to each other. In cases where they are immersed in the program or modality-wise shut off, e.g. by the use of headphones, they could make use of additional technological aids, such as ambient light that either describes the level of activity in the other program, as portrayed in Figure 1, or signals the change of channels on the other screen, e.g. by swiftly showing the first minute of the new program in a small semitransparent window, e.g. in the bottom right corner of the own screen.



Figure 1: Ambient light used to indicate the excitement level of a program shown on a different screen in the same space

3.2 Same time, different space, same content

The crucial aspect here is that people share the same physical time period in front of a screen to explicitly watch the same content. This does not mean that they have to be in the same location to share the same program. An example might be that parts of a family live in Amsterdam and the other temporarily in Moscow. If people in such a situation wish to share a program, e.g. a soccer game, a telenovella, or a talk show, they want to be aware of the emotional as well as reflective state of the others. This might be achieved by providing a semi-transparent window on the screen captured by a camera in the other location that portrays visual reactions, as displayed in Figure 2 on the next page. An audio commentary can additionally be overlaid on the actual sound of the program, which is especially effective in rooms with a surround sound system, as here a special location can be selected.

In cases that they wish to interact in real time with each other on interactive content, e.g. traversing through the story space of an i-narrative, the described real-time communication methods can be used for synchronisation. For example they can use the additional audio layer to communicate preferences for choices in which direction a story should develop, or make use of indirect communication mechanisms, such as sms-messages, that directly influence the story engine. The important aspect for all type of interactions, but most importantly for the indirect ones, is that enough feedback is provided for providing the user with the feeling that he or she is in control.



Figure 2: Transparent image overlay of co-watcher in a different location.

3.3 Same time, different space, different content

This case is similar to the case described in section 3.1. The difference here is that a spatial asynchrony is introduced. Imagine again the same constellation of a group of people who are temporarily situated in two cities but still wish to see what the others are doing without necessarily sharing the same TV viewing experience. For example, while one group watches the national news the other person or group consumes an automatically produced documentary about the happenings at the other location covering the previous week. The same audio-visual interaction methods as described in section 3.2 can be used here for quick validation of content or presentation of experiences. As one of the members perceives an interactive program (the automatically generated documentary) it might be necessary to install an additional ‘interaction’ layer, which allows the augmentation through annotation of the automatically generated content to facilitate highlighting, commenting, or altering of the content for later replay by the other group. This layer can also be used for direct manipulation in form of pointing so that, in cases of questions, the particular parts under consideration can be immediately localised. Here additional input mechanisms are required that could make use of double-touch technology for the screen, pointing devices or completely separated input devices, e.g. electronic drawing boards or

speech input, that are temporarily synchronised with the incoming content stream, allowing real time annotation.

3.4 Different time, same space, same content

The interesting change in parameters in this case is the asynchrony in time. The relevant situation is that people are interested in perceiving a program on the basis of experiences of others who already saw it but who are not around at the moment to be asked about it. The asynchrony of time allows a presentational flexibility not possible in the previous cases, which all required real-time presentation and hence generation. However, the challenging aspect of this scenario is that we need real-time annotation mechanisms that allow rich augmentation of the content so that later the experience can be partially re-established. At the same time we have to invent a new iconography to describe the augmentations on the screen so that the current user is in the position to perceive the experience of the other(s) but can also skip it and follow his or her own information or entertainment needs.

The representation of ‘experience’ is a crucial aspect. As mentioned earlier, there are several layers that need to be addressed. There is the context that needs to be represented. It would be interesting to see if a person watched the program fully or in a time constraint mode. It is also helpful to know what type of actions the person had performed during the consumption and why, resulting in a capture of actions, such as fast forward or –backward, slow motion, zoom in or pause with additional explanatory audio, visual or textual material. Most importantly, it is important to get an idea about the involvement of the other person with respect to interest and emotion. This material can be collected in audio, visual or textual form but can also be based on biometric data measurements. All this type of information can facilitate the recreation of experience. Capture and alignment of this data with the original content needs to be performed in temporal synchronisation (making use of technology already described in the previous sections) whereas the final instantiation and exploitation of the annotations can be done off-line. For example, we could calculate emotional expressions from captured face data and make use of that during the presentation of the experience enhanced content, e.g. in form of ambient light or additional iconic descriptions as overlays. At the same time the system could also make use of information gathered from actions, e.g. skipping scenes or collecting additional information from the web, for enhancing the augmentation layer of the original content. In that way the system could turn a documentary about World War II into an experienced enhanced documentary where the grandparents’ experiences (those during the war as well as those made during the perception of the program) provide an additional layer of information.

The representation problem is intrinsically connected with the presentation problem [9], as ways needed to be found

so that the additional information does not disturb the own experience but still allows alteration and comparison. As a result the presentation might make use of black bars at the top or bottom of the screen to present temporally available interactive icons that support the access as performed by others. Other options might be interactive visual or audio overlays on the content itself, or divergence methods in form of story maps to facilitate easy navigation, or split screen mechanisms. With respect to the emotional aspects necessary to re-establish an experience, one can use particular features in the collected material, e.g. smiles in detected faces, or can exploit biometric data measured during the perception of the content. The challenging aspect is that the annotations can serve as raw data and thus can be also annotated. The established layers of annotation can then be exploited by the system, for examples as summaries of programs, where particular expression aspects are highlighted, depending on the user's needs.

3.5 Different time, different space, same content

The situation covered here is that people in a different location would like to experience a program that has been watched by others before or who watch an automatically generated program, based on experience-enhanced material already available, that the other side has not yet seen either. For example, people in Moscow watch the summary of last week's TV-consumption by people in Amsterdam. Another example might be the more epic documentary of the children's spring holidays. In this situation the system relies even more on the representation and presentation methods already described, as a potential human-to-human information exchange, say for example a quick chat in the kitchen or hall, is not possible. Additional methods to those described in section 3.4 are not necessary but for the situations covered by the given context they have to be fully worked out. It is also important to mention here that all sort of automatically generated material needs to be clearly identifiable as such.

3.6 Different time, different space, different content

Besides the already mentioned problems in the context of experience representation, generation and presentation it is important to realize that there will be a growing divergence in interest between people if the separation covers a long period of time. The aspect that needs to be considered here is tracing the changes in people's behaviour and preferences. This is more a question of user modelling, which can be incorporated into the information network sketched out in the previous sections 3.1 – 3.5. If systems should be able to support richer social interaction in a TV-centred environment they have to have a notion about the particular individual, i.e. interest, preference, repugnance, and contextualised watching history. The analysis of challenges of user modelling in the context of

communicative TV is, however, beyond the scope of this paper.

4. CONCLUSION

In this article we outlined in particular participatory elements of experiences that enhance the feeling of social belonging and togetherness when TV users are separated by time and space. The discussion showed that methods addressing spatial separation in real-time communication ask for direct interaction support, which to some extent reduces the reproducibility of experiences, as they will not be stored. In cases of a temporal separation between TV users, beyond the time-zone differences, the capture of an experience is an essential element. The field of experience capture is still in its infancy but further advances are required to establish communication mechanisms in TV environments as described in this article. We showed, however, that experienced-enhanced content could establish forms of intimate communications freed from time and space. The visualisation of this type of material, as well as its production requires additional research, as the cognitive as well as procedural constraints are not yet clear.

5. REFERENCES

- [1] Cubitt, S. (2002). Spreadsheets, Sitemaps and Search Engines – Why Narrative is marginal to Multimedia and Networked Communication and Why Marginality is More Vital than Universality. In *New Screen Media – Cinema/Art/Narrative*, edited by Martin Rieser/Andrea Zapp, pp. 3 – 13, BFI Publishing, London.
- [2] Brooks KM (1999). *Metalinear Cinematic Narrative: Theory, Process, and Tool*. MIT Ph.D. Thesis.
- [3] Ursu, M.F., Kegel, I., Williams, D., Thomas, M., Mayer, H., Zsombori, V., Tuomola M.L., Larsson, H., Wyver, J., and Zimmer, R., (2008), *ShapeShifting TV - Interactive Programmes*, ACM/Springer Multimedia Systems Journal, special issue on Interactive Television, Vol. 14 (2), pp. 115–132, 2008
- [4] Stefano B. (2006): *Vox Populi: Semantic annotations and rhetorical editing for automatic video documentary generation*. University of Amsterdam Ph.D. Thesis
- [5] Davenport, G. and Murtaugh, M. (1995). *ConText towards the evolving documentary*. In *Proceedings of the Third ACM international Conference on Multimedia*, pp. 381 – 389, San Francisco, California, United States, November 05 - 09, 1995. ACM, New York, NY.
- [6] Harrison, C., Amento, B. (2007). *CollaboraTV: Using Asynchronous Communication to Make TV Social Again*. In *Interactive TV: A Shared Experience*, Adjunct Proceedings of EuroITV. Amsterdam. 218–222.
- [7] Terminal Time: <http://www.terminaltime.com/>
- [8] Nack, F. (2003) *Capturing experience - a matter of contextualising events*. Proceedings of the 1st ACM MM WS on Experiential Telepresence (ETP 03), Berkeley, CA, USA, November 7th, 2003, pp. 53 - 64
- [9] Douglas J.Y. and Hargadon, A (2004). *The Pleasures of Immersion and Interaction: Schemas, Scripts and the Fifth Business*. In *First Person – new Media as Story, Performance, and game*, ed. Noha Wardrip-Furin and Pat Harrigan, pp. 192 – 206, The MIT Press, London, England.

Strengthen the TV Experience beyond the TV

Sara Kepplinger, Florian Förster, Judith Igelsböck, Marianna Obrist

HCI & Usability Unit, ICT&S Center, University of Salzburg

Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria

sarakepplinger@gmx.at; {florian.foerster2; judith.igelsboeck; marianna.obrist}@sbg.ac.at

ABSTRACT

The TV experience is no longer restricted to the TV. Other media devices, such as the PC and mobile phones, have become platforms for consuming TV content, which are more and more extended with new features for communication and social networking. A main challenge is still how to provide users the “right” TV experience. It has to be thought about how it is possible to strengthen the TV experience including all these new opportunities. The contribution provided in this paper is based on results derived from user evaluation performed studies with social TV applications. On the basis of two case examples we outline initial assumptions on how to enable a TV experience when other devices like the PC are used beyond but in conjunction with the TV set.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems—Human factors; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented and virtual realities; H5.2 User interfaces: Theory and methods.

General Terms

Documentation, Design, Human Factors, Theory

Keywords

TV, PC, IPTV, Media Disruption, User Experience

1. INTRODUCTION & RELATED WORK

Social TV is defined as opportunity to interlink people and provide communication features to create connectedness [2], [11], [18] via the TV. Already Wellens [19] stated that “interactive television represents means of linking individuals together by providing each with an electronically mediated representation of the other’s voice and visual presence”.

A lot of applications allow social communication by using the TV set itself (e.g. SAMBA [12], Amigo TV [3], CollaboraTV [6], and Telebuddies [10]). However, we are not aware of any application for social communication on TV, which entered mainstream. These are all additional services which are not directly interlinked with the provided TV content.

But as Ursu et al. [17] state, “Television should do more than simply offer interactive services alongside (and separately from)

traditional linear programs, in the context of its dominance being seriously challenged and threatened by interactive forms of screen media entertainment” [17]. They therefore suggest the development of interactive forms of narratives as a path for the future of TV. Goldberg takes the same line by introducing the concept of semantic connection to investigate the linkage between a broadcast program and interactive television (iTV) applications [5]. To maintain the user’s immersion into a program a high semantic connection is essential. We think that these notions are equally valid for social TV – especially when other devices are used in combination with the ordinary TV [9] set in order to offer the possibility for social communication as well as for a more active participation (e.g. user-generated content [8], Messenger Zync [15]).

But the usage of additional, complementary devices leads to a media disruption. Media disruption is understood as the situation when additional effort by the user is needed (i.e. changing the device) in order to reach a certain goal [7] which is often perceived as “disruptive”. The taken effort is accepted by the user if this brings an advantage (e.g. more information, a feeling of togetherness or a more exciting TV experience). The interconnections between the TV and other media devices therefore need to be actively supported in order to strengthen the TV experience. This can be done by exploiting the advantages of from both, the TV set and other media devices, such as the PC.

However, the challenge is to maintain the TV experience when bridging different devices and to overcome and reduce the experience of media disruption. TV experience should be provided via linking content and additional services seamlessly, realizing convergence, and supporting the user to stay in the actual state of perception without a break. Within this paper we provide some answers to the question of how to strengthen the linkage between the TV and other technologies which additionally enable social communication.

2. BEYOND THE TV SET

A framework connecting the TV content and the other media services (e.g. Internet) offers a more appealing experience for the users than single services, such as TV content on the one side and communication tool on the other side.

In this section we present social TV applications we have evaluated, considering the media disruption as influence on the users’ experience of the system. Example “A” is concerned with a TV programme and the PC and example “B” addresses the connection between an IPTV portal and the PC. We will point out suggestions on how to sustain the TV experience when, for instance, the TV set and remote control are replaced by a PC and keyboard.

Evaluation results show that participants asked for a better connection between the TV and the appending technology.

Moreover, users do not mind using a combination of two different devices as long as they are linked in some way and that a mental framework (e.g. a story) is provided.

2.1 Example A: TV & PC

This example is based on an application for Augmented Virtual TV, which connects the TV content with the PC in a virtual game environment [13]. This application was evaluated as part of a national project. The provided service is a hybrid between a TV show and a game. Users meet on a virtual platform accessed via the PC and have the possibility to interact via different means of communication. They can access broadcasted TV content on demand and are offered games that allow an immersion in the broadcasted content of the TV show and to compete against each other.

Participants of a first user evaluation criticized that the prototype does not fully make use of the advantages of the TV. They suggested to couple different game levels to the progress of the TV show. Thus the time-based regularity of the TV could fuel the connection between TV show and virtual world. They furthermore asked for an overall story to strengthen the interconnection between the TV and the PC. Together they should add up to an interesting storyline around the application. Participants embraced the possibilities of social communication. However, they did not like to renounce the traditional TV experience.

2.2 Example B: IPTV & PC

In our second example, an IPTV service was provided to the users of a local community testbed, extending the known user TV experience by offering user-generated content (UGC) features on the TV combined with the PC as upload platform for audiovisual (A/V) content [14]. Additionally, a mobile phone for picture upload was provided to the users. The IPTV service was analyzed as part of the European CITIZEN MEDIA project, offering additional results from a user evaluation on User Experience and User Acceptance.

In a first development phase, people were able to produce A/V content themselves, to upload and watch content on the TV in their homes, enabled by a set-top-box (STB). To upload the produced pictures and videos, an Internet terminal and a PC were required. In a further stage a gaming approach was introduced to increase users' participation in creating and sharing content, firstly on the PC and secondly on the TV itself. Most activities happened in the PC context, but the game was extended towards the IPTV platform, and could thereby further strengthen the initial provided TV services within the testbed.

Participants of the field evaluation within this project felt sorrow about the need of an additional device (i.e. PC) for a wholly usage of the provided service and complaint about the slow response time of the TV set. A motivation is needed to keep the user being active and to prod to participation with user generated content. The majority pointed out, that their generally TV usage did not change because of the additional offer.

3. STRENGTHENING THE TV EXPERIENCE

Based on the described evaluation activities we have extracted three assumptions on how strengthen the interconnection between the TV and other media devices and how to provide the user a mental framework for keeping the TV experience beyond the TV.

3.1 Align with the Traditional TV Elements

As people define the TV experience by traditional elements as the time-schedule and storytelling, social communication application should embed these elements. They can trigger usage and bind the user to the offer. From games we can learn that they use TV story elements to offer a better experience [1]. As shown in example "A" knowledge of the TV content should be required for playing the game in order to strengthen the interconnection.

3.2 Contextualize social communication

TV is a one to many communication medium, the audience is consuming the provided content and not participating in an active way. However, interactive services and social communication in the TV context need contribution, visibility of social action and a feeling of continuation.

The professionals should provide a link back to the TV, for example through a certain focus (e.g. on a storyline, on a topic, or on a person/anchor man). It should serve as a basis for a group of people, visualize a collective progress, and provide a common experience. One example would be to enter a virtual world and chat/exchange about a common topic extracted from the TV program (see example A).

3.3 Consider User-generated content

When addressing the TV context, UGC has become relevant over the last years (e.g. having YouTube videos on the TV etc.). Providing users a possibility to actively contribute to the traditional TV content could be another opportunity for supporting the TV experience beyond the TV. The described application in Example B offered a common topic and provided a game approach to engage users. However, to keep the relation to the TV, the inclusion of UGC requires more. To make a service lively over time it needs the creation of a dramatic composition, excitement, and dates (make appointments with the user/audience) to create an entire user experience.

4. CONCLUSIONS

When additional communication devices are used for social TV, measures need to be taken not to lose the connection to the TV. The evaluation of a combination of TV, PC and game on the one hand and a combination of IPTV and PC on the other hand show that users still define their TV experience by traditional elements and do not like to miss them in the usage of social, interactive TV as well. Therefore social TV application should provide content that is related to the TV subjects offered in a scheduled form.

The question on how to strengthen the connection of TV with other technologies by facilitating social communication in order to ensure positive TV experiences is addressed within this paper by discussing three issues revealed from performed user-centered evaluation activities.

In future research and within the workshop on social communication in the TV context we want to discuss further methods and ideas towards sustaining a TV experience that is going beyond the TV set.

ACKNOWLEDGMENTS

Our thanks go to all participants of our evaluation studies, providing useful feedback, as well to all our project partners of the CITIZEN MEDIA research project (funded by FP6-2005-IST-41).

REFERENCES

- [1] Atkins, B. 2003. More than a game, the computer game as a fictional form. Manchester University Press, UK.
- [2] Cesar, P.; Chorianopoulos, K. and Jensen, J. F. Social television and user interaction Computer Entertainment, ACM, 2008, 6, 1-10.
- [3] Coppens, T., Vanparijs, F. and Handekyn, K. 2005. AmigoTV: A social TV experience through triple-play convergence. Alcatel, White Paper. Amigo TV Alcatel-Lucent
- [4] Gawlinski, M. 2003. Interactive television production. Oxford, UK, Focal Press.
- [5] Goldenberg, S., (2008). Creating Augmented and Immersive Television Experiences Using a Semantic Framework. Proceedings of the UXTV '08. P. 45-48.
- [6] Harrison, C. and Amento, B. (2007). CollaboraTV: Using Asynchronous Communication to Make TV Social Again. In: Proceedings of the EuroITV2007.
- [7] Hauff, M. (1998). Media@uni-multi.media?: Entwicklung-Gestaltung- Evaluation neuer Medien. Germany, Waxmann Verlag, p 89.
- [8] Jensen, J. F. 2007. User-generated Content – a Mega-trend in the New Media Landscape. In Adjunct Proc. of EuroITV2007, 29–30.
- [9] Klym, N., Montpetit, M., J., VCDWG, MIT CFP (2008). Innovation at the Edge: Social TV and Beyond. MIT CFP – VCDWG Working Papers, September 1, 2008, pp. 4-11.
- [10] Luyten, K., Thys, K., Huypens, S. and Connix, K. 2006. Telebuddies: Social stitching with interactive television. In Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems, 1049–1054 Telebuddies (<http://research.edm.uhasselt.be/~kris/research/projects/telebuddies/>)
- [11] Mantzari, E., Lekakos, G. and Vrechopoulos A. (2008). Social TV: Introducing Virtual Socialization in the TV Experience. Proceedings of the UXTV '08, p. 81-86.
- [12] Mayora-Ibarra, O., Costa, C. (2007). The SAMBA Approach to Community-Based Interactive Digital Television. In: Proceedings of CHINACOM '07 Second International Conference on Communications and Networking in China, 2007, pp. 676-681.
- [13] Obrist, M., Igelsböck, J., Förster, F., Tscheligi, M., Sturm, P., Hofstätter, J. And Kranzer, J. (2009). Augmented TV: Experiencing Augmented Virtual TV in a Game-based PC Environment. Accepted poster for EuroITV2009 Conference, Leuven, Belgium, 3-5 June, 2009.
- [14] Obrist, M., Miletich, M., Holocher, T., Beck, E., Kepplinger, S., Muzak, P., Bernhaupt, R., Tscheligi, M. (accepted). Local Communities & IPTV: Lessons Learned in an Early Design and Development Phase. In: Computers in Entertainment. To appear 2009.
- [15] Shamma D., Bastéa-Forte M., Joubert N. and Liu Y. (2008). Enhancing Online Personal Connections through the Synchronized Sharing of Online Video. Proceedings of CHI 2008, p 2931-2936.
- [16] Svensson, M. S. and Sokoler, T. (2008). Ticket-to-Talk-Television: Designing for the circumstantial nature of everyday social interaction. Proc. of NordiCHI 2008.
- [17] Ursu, M., F., Thomas, M., Kegel, I., Williams, D., Tuomola, M., Lindstedt, I., Wright, T., Leuridijk, A., Zsombori, V., Sussner, J., Myrestam, U. and Hall, N., (2008). Interactive TV Narratives: Opportunities, Progress, and Challenges. ACM Transactions on Multimedia Computing, Communications and Applications, Vol.4, No. 4, Article 25, October 2008, ACM Press, New York, USA.
- [18] Weisz, J. D., Kiesler S., Zhang H., Ren Y., Kraut R. E. and Konstan J. A. (2007). Watching Together: Integrating Text Chat with Video. In proceedings of CHI 2007. p. 877-886.
- [19] Wellens, A., R. (1979). An interactive television laboratory for the study of social interaction. In: Journal of Nonverbal Behavior, Volume 4 / Number 2, December 1979, Springer Netherlands, pp. 119-122.

Sharing Enriched Interactive TV Experiences with the iNEM4U Software Framework

Cristian Hesselman
Novay, The Netherlands
cristian.hesselman@novay.nl

Keir Shepherd
HP Labs, UK
keir.shepherd@hp.com

ABSTRACT

The project iNEM4U (Interactive Networked Experiences in Multimedia for You) is developing an open distributed software framework that allows users and service providers to seamlessly combine interactive multimedia content and services from different types of networks (the Internet, in-home, mobile, and IPTV networks) into one shared experience. In this paper, we provide a brief overview of the project and discuss the TV-centric application we are using to enhance people's social media experiences, in particular to increase the feeling of connectedness for groups of people, such as families and friends.

Categories and Subject Descriptors

C.2.4 [Computer-Communication Networks]: Distributed Systems – *distributed applications*.

General Terms

Design

Keywords

Enriched social multimedia experiences, interactive TV, cross-technology software framework

1. INTRODUCTION

Today's consumers are typically at the center of different types of network infrastructures, such as TV networks, the Internet, and in-home and mobile networks. Each of these networks has its own unique characteristics in terms of supported content types, services, interactivity levels, and evolution speeds. For example, IPTV networks usually provide high quality professional content with remote control-based interactivity, whereas the Internet typically offers services that provide low quality user-generated content with high levels of interactivity. Another example is that IP-based networks such as the Internet and IPTV networks are much more versatile, in particular supporting bi-directional communication, than traditional networks such as terrestrial and satellite networks and also evolve much more quickly.

In iNEM4U we believe that we can facilitate a new generation of social media experiences by enabling users and service providers to combine services, content sources, and devices from *different* networks into one "sharing session", thus

mixing and exploiting the unique characteristics of the individual networks. The problem, however, is that it is currently not possible for users and service providers to accomplish this in an easy way. For example, it is impossible for an average user to set up the facilities to get a live feed with high quality content from a rock concert on his TV set (IPTV network), have his buddies on the Internet get the same content in a synchronized way, enable the whole group to get live picture-in-picture overlays showing photos that other friends who are actually at the concert took with their cell phones (mobile network), and after the concert buy a HD-quality video of the concert on the Web using their mobile account. This also hinders the materialization of new business models and service concepts that span multiple network infrastructures, for instance because providers cannot efficiently make their services available across different types of networks.

The goal of iNEM4U is to overcome these problems by means of a software infrastructure that allows both users and providers to seamlessly combine multimedia content, value-added services, and devices from different networks into one interactive multi-user "session". In iNEM4U, we refer to these networks as "technology domains", as the root cause of the problem outlined above is that each network uses its own set technologies, for instance in terms of application-level protocols, user identities, group descriptions, and metadata formats.

2. SCENARIO

Figure 1 shows the scenario the iNEM4U project is using as its key use case [1]. The scenario revolves around a socially augmented event (a concert to be specific) in which four participants, two who are in the concert audience (Harry and Pauline) and two who are watching the event on TVs remotely

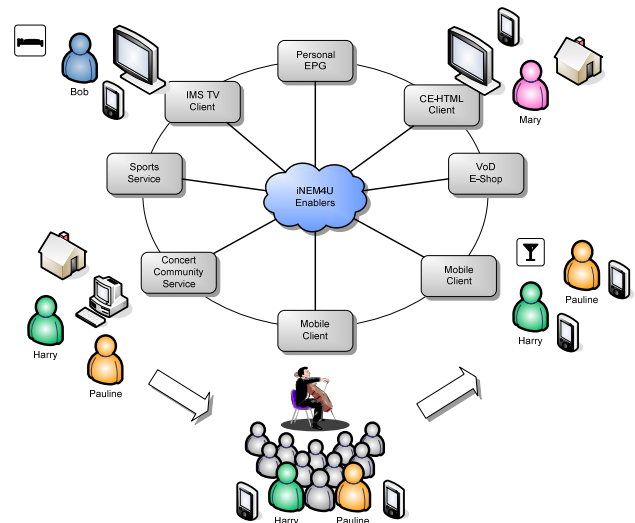


Figure 1. iNEM4U scenario.

(Bob and Mary), enrich the group's experience through social interaction and media sharing. Figure 1 shows the applications involved and the iNEM4U platform in the cloud in the middle.

The scenario begins with the creation of a dynamic community (known as an ad-hoc group). Through the mining of buddy lists and context the iNEM4U platform enables an ad-hoc group, that exists for the duration of the concert, to be created based on friends who are attending the concert, either physically (Harry and Pauline) or remotely, watching a live broadcast (Bob and Mary).

As the four friends join the ad-hoc group through their various devices an iNEM4U sharing session is created through which media and interactions can be shared, synchronized, adapted, recorded and played back. As the concert progresses Harry and Pauline use their mobile phones to take pictures and videos and add these to the iNEM4U session. Harry and Pauline's content is rendered as picture-in-picture overlays for Bob and Mary on their TVs giving them a view of the concert from the audiences point of view.

During the break Bob browses the video archive catalogue for the artist using an interactive TV application and adds some interesting VoD content to the session. Mary, who is at home, has left the room briefly so the content does not begin to play on her TV until she returns.

Following the concert, and with the consent of the four friends, the session information is analysed and a special group offer of access to VoD content from other artists they may like is made to them as part of the iNEM4U session. If all four choose to take it then they will be given a significant discount.

3. APPROACH

Figure 2 shows the approach that iNEM4U takes to realize scenarios such as the one outlined in Section 2. Our technical work focuses on R&D on the iNEM4U software framework, which allows social applications to easily combine the unique content sources and services from different technology domains (home, broadcast, mobile, Web) into one cross-domain iNEM4U session, for instance to enrich social interactive TV applications. The novelty of the iNEM4U framework is that it consists of a set

of service enablers that provide *rich* functionalities and APIs for applications to manage iNEM4U sharing sessions. Examples are enablers for cross-domain identity management, cross-domain synchronization of content at different locations, and cross-domain community representations. The iNEM4U platform is also open, which, combined with its rich enabling services, sets it apart from prior work in this area, such as [2] and [3].

The project also investigates novel business models enabled by the iNEM4U infrastructure, which are discussed in [4].

4. DISCUSSION

The iNEM4U framework delivers a rich set of enablers which through a variety of techniques, including context and metadata derived communities, multimedia shared interactive sessions, synchronized media consumption, and co-browsing allows service providers to create and deliver timely and socially-enriched multimedia experiences involving a mix of devices, user-generated content, professional content, and value-added services. We expect that this will facilitate much richer shared experiences than possible today, for instance for social TV applications, thus enabling geographically dispersed groups of people such as families and friends to maintain a stronger feeling of togetherness.

However, togetherness is not simply fostered through rich multimedia experiences; it also relies considerably on being able to deliver the right thing in the moment. The iNEM4U platform is ideally suited for this purpose as its cross-domain enablers allow shared interactive multimedia experiences, of the type discussed in Section 2, to be delivered to communities no matter where the individuals are and how they are connected. Based on our current prototype, we believe that by bringing people together through a synchronised multimedia sessions across mobile (3G handset) and fixed (TV) devices we can deliver an overall experience which is more rewarding for all participants than watching the content alone and can even improve the subjective perception of the content, with encoding and even content quality issues becoming less distracting when watched with friends.

We will test this hypothesis and evaluate the impact of cross-domain sharing sessions on people's feeling of connectedness using more objective studies carried out in a representative lab environment at the end of the project.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216647. The project's website is at www.inem4u.eu.

6. REFERENCES

- [1] iNEM4U deliverable D5.1, "Demo Scenarios", <https://doc.telin.nl/dsweb/Get/Document-95058>, Dec 2008
- [2] R. Schatz, S. Wagner, S. Egger, and N. Jordan, "Mobile TV becomes Social: Integrating Content with Communications", ITI 2007, Cavtat, Croatia, June 2007
- [3] M. Baca and H. Holtzman, "Television meets Facebook: Social Networks through Consumer Electronics", Workshop on Sharing Content and Experiences with Social Interactive Television, EuroITV2008, Salzburg, Austria, July 2008
- [4] iNEM4U deliverable D2.3, "Description of the Selected Business Model", <https://doc.telin.nl/dsweb/Get/Document-95057>, Dec 2008

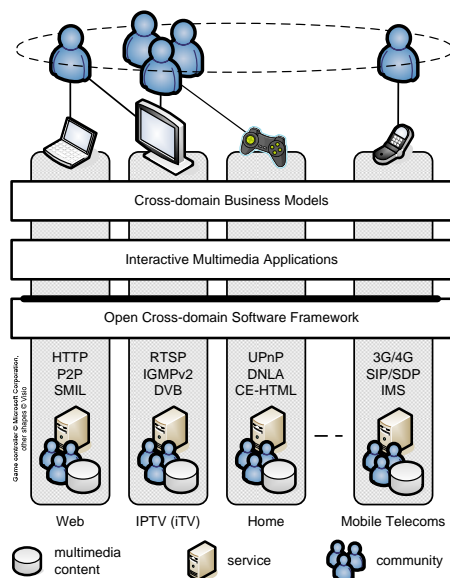


Figure 2. iNEM4U approach.

From Social TV to Structured Communication Formats

Gunnar Harboe

Motorola

1295 East Algonquin Road

Schaumburg, IL 60196

gunnar.harboe@motorola.com

Santosh Basapur

Motorola

1295 East Algonquin Road

Schaumburg, IL 60196

sbasapur@motorola.com

ABSTRACT

We describe research and ideas that could inform the design of social experiences around the TV. This work is primarily based in the field of social television, which we summarize briefly. We focus on the idea of structuring and interpreting freeform communication so that it can be used to trigger system actions, which in turn can provide distinct social experiences. We first analyze social program recommendations from this perspective, then look at a new genre of social video game that blurs text chat and command input.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors.

Keywords

Social television, interactive television, communication, CMC, field studies, social games, adventure games.

1. INTRODUCTION

For the past decade there has been substantial research into *social television*, defined as social experiences around television content, and systems to support these experiences (usually remotely). Social TV systems typically integrate presence awareness and some form of communication (text, voice or video chat) with shared viewing of video content. The most current findings and achievements in the field of social television are described by Cesar, Geerts and Chorianopolous [3].

In previous research we have shown that the capability for users to communicate freely in their own words is critical to a rewarding Social TV experience [4]. At the same time, there is growing recognition of the potential found in more “intelligent” communication features that will allow users to engage in shared activities. A simple and widespread example of this kind of feature is the program invitation, recommendation or suggestion,

which lets a user tune directly to a particular show and makes it easy for two buddies to synchronize their viewing.

One promising possibility is the creation of social experiences tailored to specific TV programs, such as multiplayer competitions within quiz shows, fact checking or commentary during political debates, or a gossip channel as part of entertainment news. Enabling third-party creation of these custom experiences requires standardized protocols and an API for the intelligent communication features. The ability to specify different communication formats in a consistent way could lead to the invention of a wide variety of TV-centered structured social experiences. To design these experiences, we need to understand the users and their needs, but we also need creativity and vision.

The following sections first provide an overview of relevant TV-centered communication formats found in the literature and in commercial products, then describe findings from two user studies that looked at program recommendations. Finally we jump beyond the data and propose an idea for a novel TV-centered social game.

2. RELATED WORK

Besides program recommendations, which offer an implicit or explicit choice to accept or ignore the suggestion, several systems have offered other forms of messages with multiple-choice replies. 2BeOn let users send ClipEmails, quick questions that allow only yes/no answers, to their buddies [1]. In ConnetTV a similar feature allowed the operators to gather experience samples from users, who were presented with multiple-choice questions at particular times in the trial [2]. It is easy to see how these capabilities could be applied to classic Interactive TV use-cases like voting and surveys.

Telebuddies allows programs to be annotated with an interaction script that for example can define a multi-player quiz show running on top of the show [6]. Staged Participatory Multimedia Events (SPMEs) extend this idea to creating fully user-generated experiences, such as a live online auction through the TV [9].

On the product side, companies are also working to create tailored social experiences and activities. Disney BD-Live offers multiplayer quiz games on certain Blu-ray discs, based on clips from the movie.¹ And MTV Backchannel allows players to add sarcastic or witty comments to reality shows like *The Hills* and *Real World*, an interaction format that is particularly suited to that program genre.²

¹ <http://www.disneybdlivenetwork.com/>

² <http://backchannel.mtv.com/>

3. USER RESEARCH FINDINGS

3.1 Suggestions in Social TV

We have done several user studies that looked at program recommendations (suggestions). In the first user trial we provided a system that offered presence, program suggestions and closed-form messages (emoticons, a list of text messages to choose from), but no free-form communication [7]. Participants liked the ability to send suggestions and used it often (178 times by 10 households over 2 weeks). They also didn't mind receiving them. As one participant said: "She watches the Food network a lot. She's given me some funny suggestions for things to watch, where I normally wouldn't tune in, so I've taken her suggestions and I've gone to them." However, only 28 of the suggestions (15.7%) were in fact accepted.

Several participants were frustrated by their inability to convince others to accept their suggestions. One told us: "I have bumped over to their shows so I can talk to them and say, 'Oh come on, change to my show!' But they don't budge." They ascribed this to the restrictive nature of the suggestions: "You can't really specify 'Great show!'" If there was a way to explain *why* they might want to watch something, they might be more easily persuaded.

In our subsequent trials ([8], [5]) we provided freeform communication through text chat and voice calls, without making any changes to the suggestions. Participants were no longer frustrated by an inability to express what they wanted to communicate, but use of the suggestions dropped precipitously. In the last of the trials, the 5 households in the trial sent only 11 suggestions over 2 weeks. Perhaps suggestions, with their sparse and impersonal information, were made redundant by the ability to chat freely. Nevertheless, we persist in believing that the basic idea is useful.

3.2 Actual Recommendation Practices

In a recent study, we looked at people's actual practices around TV content and interactive television features, including how they give and receive recommendations to others, and how they respond to system-generated recommendations. Six participants were interviewed in their homes using a semi-structured interview protocol, and key statements extracted and clustered in an affinity diagram, with the following findings emerging from the patterns in the data.

Several participants resisted the idea of choosing what to watch based on recommendations (or reviews and ratings), whether from friends, critics, random strangers or computer algorithms. As one put it: "I like to check myself. I like to be my own judge." They were often not interested in whether something had critical or popular approval, seeing media content preferences as wholly subjective: "I know what they're like, so I don't really care what everybody else rates it."

At the same time, we know that no one watches or even samples everything, so clearly they are making choices based on the information they have. Our participants acknowledged this. Even in denying looking at recommendations, one explained it was "because usually I've heard about it on TV or something beforehand." Another said, "If something got one star and it sounds pretty interesting, I'll still [...] watch it."

Participants didn't pay too much attention to recommendations because they didn't find them reliable. Even friends who liked many of the same things would often give them recommendations to things they ended up not liking, and automatic recommendations did worse still: "What it usually picks up is older shows that I used to watch [...] It picks up a ton of reruns." Several participants had DVR systems that could auto-record suggestions, but all had turned that feature off.

At the same time, many participants still enjoyed receiving recommendations: "I *really* like when people recommend things. I like to hear their opinions." The explanation for this seems to be that sharing recommendations is not primarily about finding something to watch, but a way to socialize. Recommendations come up spontaneously in the context of conversations. "We were talking about a show that we both watch [*True Blood*] before *Entourage*. And I said, 'Oh, you still haven't caught *Entourage*?' She said no. 'You should check it out!'" They also help build relationships by providing common ground for conversation and experience sharing: "We were talking about, you know, could there be UFOs, you know, flying in? [...] So I was just saying, when you're on the History Channel, just look for the USO [Unidentified Submerged Object] shows... documentaries, and they'll inform you."

Because a successful suggestion can help create common ground and strengthen the relationship, the person giving the suggestion often tries to make it as persuasive as possible. To that end, they would describe it, give reasons to watch it, even screen short clips. "To my other friend [...] I would say, 'Hey, have you watched *The Lost* [*sic*]? It's a good show! So-and-so is pretty hot.' You know: 'There is this one girl...' Whatever reason they would accept [as an] attraction to watch that show."

These findings suggest that social recommendations have the potential to be more appealing than system recommendations ever could, but that they need to be integrated organically in conversations, and should not necessarily be evaluated on their "click-through" rate. We continue to explore designs that acknowledge recommending as an inherently social activity.

4. SOCIAL ADVENTURE GAMES

Integrating interactive recommendations into natural conversations is one very simple attempt to support a more structured activity (in this case recording and watching TV programs) within the context of freeform communication. We can take this idea much further: The Social TV setup from our trials, with a wireless QWERTY keyboard for convenient text entry, provides a promising platform for creating other TV-centered social experiences. Here we outline an idea for a novel social game based around text chatting.

Adventure games, where players solve puzzles to advance a story, were one of the most popular computer game genres in the 1980s and early 1990s, first text adventures such as *Zork* (1980) and *The Hitchhiker's Guide to the Galaxy* (1984), later graphic adventures such as *King's Quest* (1984) and *The Secret of Monkey Island* (1990). Since then, they have been eclipsed by other types of games. The reasons proposed for this are many, but two are worth mentioning: First, the genre failed to adapt to the rise of the Internet and multiplayer gaming. There have been no successful multiplayer adventures. Secondly, neither of the genre's two main

control paradigms: text parser and point-and-click, were suitable for videogame consoles. (until the Wii console introduced a pointer).

This game gets around those problems by embedding a text parser in a text chat. In games controlled by text parser, the player types commands into a command line (e.g. “look at table,” “pick up book”), which are then interpreted and performed within the game. In this new version of the idea, all players share a view of the game world, and can chat with each other via text messages. If the message is recognized as a command, it becomes available to perform within the game.

One possible design would be to execute all valid commands, but this could easily become confusing and chaotic. It would probably be better to mimic the situation where multiple people are huddled around a game, offering suggestions to the person at the controls. In other words, one of the players is responsible for selecting which recognized commands, typed by the other players, to actually perform. If that player is also blocked from chatting or entering commands, then cooperation is required in order to communicate and find the right sequence of actions to advance in the game. At the same time, individual players could gain points based on their contribution to the solution, setting up competition between them (Figure 1).

Unlike other multiplayer games, where each player controls a different character or team, this idea shares control of one character between multiple players, who all view the same thing. For that reason, and because adventure games have already developed interactive storytelling and cinematic techniques to a high level, this new game type would be better placed to take advantage of television’s narrative forms.

5. CONCLUSION

The two specific features we have focused on in this paper, recommendations and multiplayer adventure games, may seem unrelated, but what they have in common is the view of communication and shared content as a platform for more structured interactions. We see this as a promising direction for TV-centered social experiences.

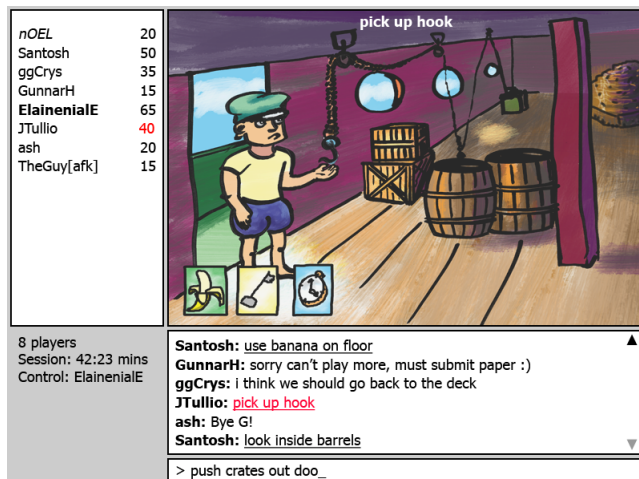


Figure 1. Sketch of a social adventure game that combines a text parser with freeform chat. Messages that are recognized as commands become available to be performed in the game.

6. ACKNOWLEDGMENTS

Our thanks to Ashley Novak, Hiren Mandalia, Crysta Metcalf, Noel Massey, Frank Bentley, Elaine Huang, Joe Tullio and Guy Romano for their help and support.

7. REFERENCES

- [1] Abreu, J., Almeida, P., and Branco, V. 2001. 2BeOn: Interactive Television Supporting Interpersonal Communication. In J. A. Jorge, N. Correia, H. Jones, and M. B. Kamegai, (Eds.), *Proceedings of the Sixth Eurographics Workshop on Multimedia 2001*. Springer Verlag, New York, NY, 199–208.
- [2] Boertjes, E., Klok, J., Niamut, O., and Staal, M. 2009. ConnetTV: Share the Experience. In P. Cesar, D. Geerts, and K. Chorianopolous (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Reference, Hershey, PA, 189–203.
- [3] Cesar, P., Geerts, D., and Chorianopolous, K. (Eds.) 2009. *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Reference, Hershey, PA.
- [4] Harboe, G., Huang, E., Massey, N., Metcalf, C., Novak, A., Romano, G., and Tullio, J. 2009. Getting to Know Social Television: One Team’s Discoveries from Library to Living Room. In P. Cesar, D. Geerts, and K. Chorianopolous (Eds.), *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Reference, Hershey, PA, 159–188.
- [5] Huang, E., Harboe, G., Tullio, J., Novak, A., Massey, N., Metcalf, C.J., and Romano, G. 2009. Social Television Comes Home: A Field Study of Communication Choices and Practices in TV-Based Text and Voice Chat. In *Proceeding of the Twenty-Seventh Annual SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, NY.
- [6] Luyten, K., Thys, K., Huypens, S., and Coninx, K. 2006. Telebuddies: Social Stitching with Interactive Television. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, New York, NY, 1049–1054.
- [7] Metcalf, C., Harboe, G., Massey, N., Tullio, J., Romano, G., Huang, E.M., and Bentley, F. 2008. Examining Presence and Lightweight Messaging in a Social Television Experience. *ACM Trans. Multimedia Comput. Commun. Appl.* 4, 4 (Oct. 2008), 1–16.
- [8] Tullio, J., Harboe, G., and Massey, N. 2008. Investigating the Use of Voice and Text Chat in a Social Television System. In M. Tscheligi, M. Obrist, and A. Lugmayr (Eds.) *Proceedings of the 6th European Conference, EuroITV 2008*. Springer, Berlin, 163–167.
- [9] Van den Bergh, J., et al. 2007. Model-Driven Creation of Staged Participatory Multimedia Events on TV. In P. Cesar et al. (Eds.), *Proceedings of the 5th European Conference, EuroITV 2007*. Springer, Berlin, 21–30.

Presence and mediated interaction: A means to an end?

Lizzy Bleumers
IBBT-SMIT VUB

Pleinlaan 2
1050 Brussels, Belgium
+32 2 629 16 47

lizzy.bleumers@vub.ac.be

Tim Van Lier
IBBT-SMIT VUB

Pleinlaan 2
1050 Brussels, Belgium
+32 2 629 16 35

tivlier@vub.ac.be

An Jacobs
IBBT-SMIT VUB

Pleinlaan 2
1050 Brussels, Belgium
+32 2 629 16 48

an.jacobs@vub.ac.be

ABSTRACT

Promoting a sense of presence is often identified as a prerequisite for mediated interaction. To do so, however, we need a thorough understanding of what presence encompasses and how it can be influenced. The goal of this paper is to elaborate on the different aspects of the sense of presence as identified in the literature, while illustrating whether and how these aspects are promoted in three virtual world cases. We hope to evoke reflection on the link between promoting presence and supporting mediated interaction.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces]: Theory and models

General Terms

Design, Theory.

Keywords

Presence, Mediated interaction, Virtual worlds, Case studies

1. INTRODUCTION

Bronack et al. [2] identify promotion of presence as an essential component of an effectively designed virtual world, that is, a virtual world that enables effortless meaningful communication between and towards its participants. To promote the sense of presence - in the design of virtual worlds and mediated interaction applications in general - we need clarity on what it comprises and what factors determine it.

The complex nature of presence becomes apparent in the statement on this concept issued by The International Society of Presence Research [6]. Summarizing this statement, presence can be described as a variable psychological state or subjective perception that can be part of any technology-mediated experience, that is influenced by factors proper to the medium, the represented content and the user, and that is multi-dimensional.

In our recent work, we consulted the literature on presence (specifically, [1], [4], [5], [8] and [9]) to gain insight on the

various dimensions of presence. From these reference works, we extracted a hypothetical framework that attempts to link the various dimensions of presence and their possible determinants. We then applied this framework to three cases.

The selected cases were three virtual worlds for a young audience: Neopets (www.neopets.com), Chobots (www.chobots.com) and UBFunkeys (www.ubfunkeys.com). Neopets' website offers a virtual world/community where users become the owners of one or more virtual pets. Chobots is a web application offering a family oriented virtual world in which users run about as aliens. UBFunkeys is a software application that comes with tangible toy figures needed to access different parts of the virtual world.

The goal of this paper is to concisely present the various aspects of presence, each time illustrating how these aspects of presence are promoted or counteracted in the selected cases.

2. DIMENSIONS OF PRESENCE

The descriptions of the different aspects of presence in this paper are based primarily on the seminal work of Lombard and Ditton [8], who conducted a review of the literature on presence. Although they wrote their review twelve years ago, the dimensions of presence they identified, are still the main conceptualizations found in the literature today. We adapted their descriptions based on related works referred to in the introduction and based on personal critical arguments.

2.1 Social presence

A first conceptualization of presence that Lombard and Ditton [8] derive from the literature is that of social presence. Biocca et al. [1] identify three seemingly interrelated dimensions of social presence: co-presence, psychological involvement and behavioral engagement. Users first need to become aware that another actor is within range (i.e., co-presence) before a psychological relationship with that other may be established (i.e., psychological involvement). The latter occurs when users pick up (non)verbal cues that signal the other's intentions and thus enable mutual understanding. As a source of such cues, behavioral engagement, such as play, contributes to building psychological rapport.

In the cases we analyzed, co-presence is supported by direct and indirect evidence of other users. In Chobots, for example, the visualization of users in the virtual world as alien avatars provides direct evidence of their presence. In Neopets, the high scores listed in the mini-games implicitly suggest that other players were, and may still be, around.

In the selected cases, users can gain mutual understanding through various means of interaction. For instance, they can engage in chat

or multiplayer games as forms of synchronous verbal and non-verbal interaction. They can start a conversation on the forum or, as on Neopets, exchange items as forms of asynchronous verbal and non-verbal interaction. In addition, the cases show that users can also convey their feelings and beliefs in a non-reciprocal way, for instance, through profile pages and avatar personalization.

2.2 Realism

A second way in which presence has been conceived involves the sense of realism. According to Lombard and Ditton [8] this can refer to two types of realism: perceptual realism and plausibility (or social realism). We believe realism of interaction should be included since the first two only involve realism of representation.

2.2.1 Perceptual realism

This first component of the sense of realism corresponds to the sense that objects, events and people represented by the medium look, sound, feel, taste and smell like the “real” thing.

In the three cases, the cartoon like style of the virtual worlds counteracts the sense of perceptual realism. So does the fact that not all the user’s senses are addressed. Indeed, the cases mainly stimulate the visual and auditory senses, although UBFunkeys adds a tactile dimension to the user experience.

2.2.2 Plausibility

This second component refers to the sense that what is represented is plausible, but not necessarily “real” as described above. According to Connell and Keane [3], a sequence of events is judged to be plausible when it matches users’ prior knowledge and experience.

When our knowledge of the everyday world is taken as the frame of reference, the cases we analyzed do not offer very plausible content. The virtual worlds, its inhabitants and some of their behaviors are clearly fictional. Still, there are parallels with everyday life such as the performance of routine activities (e.g., eating).

2.2.3 Realism of interaction

A third component of the sense of realism is the sense that the way we interact with the objects and people represented by the medium is similar to real world interaction.

The selected cases involve conventional computer interaction, which bears little resemblance with real world interaction. Users navigate and select items using the computer mouse and occasionally use the keyboard (e.g., for chat).

2.3 Transportation

The third conceptualization that Lombard and Ditton [8] distinguish is presence as transportation. This can manifest itself in different forms. Users can have the sense that something is brought to them, e.g., when listening to the radio. In virtual world applications, users often have the experience of being some place other than their physical location. Sometimes, they share that virtual place with other users.

The sense of transportation requires that users are able to mentally construct a space or scene and are then able to accept that they are in it. In the selected cases, the user is given a sense of space by the implementation of depth cues in the visualization of the virtual world, such as occlusion and perspective. A combination of

storyline, visuals and musical ambience helps to set the scene; the emphasis on each of these components varies from case to case.

Acceptance of being in the virtual world is believed to be facilitated by seeing (part of) yourself there [4] and by having control over it ([4],[9]). Both Chobots and UBFunkeys provide users with visual evidence of themselves, in particular, by offering them a godlike perspective (i.e., third person perspective). In Neopets, however, users do not see themselves in the virtual world, which is detrimental for the sense of transportation. In all three cases, users mainly have substantial control on their environment in the (mini-)games which are to a greater (Chobots) or lesser extent (Neopets) integrated in the virtual world.

2.4 Immersion

The fourth conceptualization of presence that Lombard and Ditton [8] identify is immersion, which has a sensory and an attentional component. As the authors explain, strong sensory immersion refers to a state in which the user’s senses are all fully addressed by the medium. As such, this component depends on the capacity of the medium to stimulate various senses and shut out outside influences. Strong attentional immersion refers to a state in which the user’s attention is entirely devoted to the mediated content. Thus, this component depends on the availability of content that is salient enough to grab and relevant enough to hold the user’s attention and on the presence of possible distracters (see the literature on bottom-up and top-down attention, e.g., [7]).

In the cases we studied, sensory immersion is not strongly promoted due to their emphasis on visual stimulation and their reliance on the user to reduce the impact of external influences such as environmental noise. While the cases provide colorful content that easily attracts attention, it is difficult to predict whether users will find the content relevant and thus reach attentional immersion. We did observe that in Neopets it is harder to focus because of the abundance of content items, relevant and irrelevant (i.e., advertisements), competing for attention.

2.5 Social actor

The fifth conceptualization of presence listed by Lombard and Ditton [8], relates to phenomena such as interacting with a computer-generated character or replying to a person who is talking on television. Common to these phenomena is that the user responds to or interacts with an entity that is not an autonomous intelligent being and/or is not able to reciprocate the user.

According to Lombard and Ditton, users fail to acknowledge that the entity is not a social actor. However, we believe that users may exhibit this behavior even though they are well aware that the entity they are dealing with is not a social actor. For instance, they may automatically respond to certain social cues (e.g., smiling to the sympathetic person on tv) or they may act with another goal than to actually interact (e.g., venting frustration). If the user indeed believes an encountered entity to be a mediated social actor just like him or her, the experience of that entity’s presence is likely to be the same as what we described in Section 2.1. If this is not the case, however, this experience is likely to differ.

While examining our three cases, repetition and lack of (or limited) interactivity appeared to be clues that a character was computer-generated. Interestingly, a character that was evidently computer-generated, a virtual pet, evoked emotional responses by expressing needs and emotions.

2.6 Medium as social actor

Lombard and Ditton [8] distinguish a sixth conceptualization of presence that relates to phenomena similar to those in 2.5. Here, the medium itself, instead of a mediated entity, is treated as a social actor. Because the focus of our cases studies was on presence within the applications and not on presence of the medium itself, we did not include it and will not discuss it further.

3. LESSONS LEARNT

While looking for potential cases and analyzing the selected cases, we have gained a number of insights that we would like to share.

We found that developers of virtual worlds for children have concerns for the risks tied to social presence, such as harassment. In the selected cases, these concerns are mirrored in the implementation of predefined or moderated chat. We also noticed applications in which access to other users was impossible. This solution seems too radical; although the risks of social presence are reduced, the benefits of social presence are also lost.

The characteristics of the selected cases do not support a realistic experience. However, the apparent success of these and similar cases compels us to nuance the importance of realism. Indeed, it may be more appropriate that the representation and interaction style match the goal and context of use. Also, users might not use knowledge of the everyday world but of the virtual world as a reference. Finally, consistent mapping may be more important for presence than “natural” mapping. This requires further research.

When studying the different cases, we found that a sense of space is evoked by means of occlusion and perspective. Furthermore, users are given evidence of themselves in the virtual world through a third person perspective (at least, in Chobots and UBFunkeys). Interestingly, these factors seem to compensate for the absence of factors more commonly related to the sense of transportation, i.e., 3D stimulation and a first person perspective.

In the cases studied, it was hard to establish whether attentional immersion is promoted, because the relevance of the content depends on the user’s interests. Developers may try to increase the probability that users find something of interest by offering a great variety of content. However, given our experience with the cases, we emphasize that the content should be provided in a structured way, lest the users lose focus. In addition, the application should be marketed appropriately so that new users are not disappointed.

Finally, we found that the virtual characters can have very different roles. Certain characters are simply intended to make the virtual world seem less empty, while others are intended to establish an affective relationship with (e.g., virtual pets). It appears to us that this role can help to determine how much effort should be made to make the character appear to be a social actor.

4. DISCUSSION

In this paper, we gave a brief overview of the various dimensions of presence based on existing literature and we illustrated how these aspects of presence are promoted or inhibited in a selection of virtual world applications appealing to a young audience.

Our examples pertained to medium and content characteristics that are believed to promote or hinder the sense of presence.

However, user and context characteristics may interact with medium and content characteristics so that they enhance, dampen or even eliminate effects of the latter. They may also compensate, to some extent, for a lack of appropriate medium and content characteristics. For instance, the imagination of the user may compensate for the lack of a strong narrative to set the scene.

To conclude, we return to Bronack et al. [2] who argue that promoting presence is essential for meaningful interaction. This begs the question of which dimensions of presence are essential for meaningful interaction? And, what degree of presence is optimal? For instance, it seems straightforward that being able to establish a psychological relationship with others will enhance interaction, but sharing a virtual space may also be beneficial for social interaction. With regard to the second question, we may be tempted to answer that maximal presence is optimal. However, too high presence of users, for example, may actually inhibit interaction. These issues merit further attention of researchers and practitioners who are interested in enhancing mediated interaction.

5. ACKNOWLEDGMENTS

This workshop paper is based on work that was done as part of the IBBT-project Hi-Masquerade. More information on the project can be found at <http://www.ibbt.be/en/project/hi-masquerade>.

6. REFERENCES

- [1] Biocca, F., Harms, C., & Burgoon, J. K. (2003). Towards a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and virtual environments*, 12(5), 456-480.
- [2] Bronack, S., Cheney, A., Riedl, R., & Tashner, J. (2008). Designing virtual worlds to facilitate communication: Issues, considerations, and lessons learned. *Technical Communication*, 55(3), 261-269.
- [3] Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science*, 30(1), 95-120.
- [4] Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1(2), 262-271.
- [5] Ijsselstein, W., A., de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: Concept, determinants and measurement. *Proceedings of the SPIE*, 3959, 520-529.
- [6] International Society for Presence Research. (2000). *The Concept of Presence: Explication Statement*. Retrieved from <http://ispr.info/>
- [7] Kim, M. S., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. *Perception and Psychophysics*, 61(6), 1009-1023.
- [8] Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication*, 3(2), Retrieved from: <http://www.ascusc.org/jcmc/vol3/issue2/lombard.html>
- [9] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4), 73-93.

Workshop 4:

Defining the Architecture for Next Generation Inclusive Television

Defining the Architecture for Next Generation Inclusive Television

Richard N. Griffiths

School of Computing, Mathematical & Information
Sciences, University of Brighton
Watts Building, Lewes Road
Moulsecoomb, Brighton, BN2 4GJ, UK
+44 273 642477

r.n.griffiths@brighton.ac.uk

Mark V. Springett

Interaction Design Centre, School of Computing and
Multimedia Technology, Middlesex University
Town Hall, The Burroughs
Hendon, London, NW4 4BT, UK
+44 208 411 5015

m.springett@mdx.ac.uk

ABSTRACT

For inclusivity and accessibility, the switchover to digital television (DTV) presents great possibilities, but has also introduced new frustrations. New architectures for television reception are in development that in principle will enable much more creative solutions. The workshop is intended to contribute to the development of these new architectures by exploring the issues and problems for inclusive design in digital television, and proposing methods, enhancements and adaptations that new architectures should enable. DTV Practitioners both with and without experience of inclusive development are invited to contribute.

Categories and Subject Descriptors

C.2 Computer-Communication Networks

K.4.2 [Social Issues]: Assistive technologies for persons with disabilities, Handicapped persons/special needs.

General Terms

Design, Economics, Human Factors, Standardization.

Keywords

Television, inclusive design, accessibility, usability, architectures.

1. INTRODUCTION

There is an unresolved paradox in the development of digital interactive television. The replacement of specialised analogue by generalised digital technology has brought about a step change in the flexibility with which broadcast services are consumed. This should imply an improvement in accessibility for audience members with impairments. In practice however, there is an increasing exclusion of segments of the audience. Previously simple (if limited in functionality) analogue television sets are replaced with impossibly complex feature rich digital ones. The introduction of a technology that should make things better is actually making them worse!

For people with visual impairments, there is an increased need to read text and interpret graphical displays. For people with motor impairments there is increased need to make fine manipulations of complex remote controls. For people with cognitive impairments there is an increased need to make sense of a complex communications technology; one by which they may enter into contractual obligations with a range of programme, service and

product providers. Some advantage has been taken of the increased functionality available to support people with no, or reduced hearing through selectable subtitles and signing, and for the blind through audio description. However this has been restricted so far to only a proportion of 'broadcast' programming. The growing resource of video on demand and user generated content is not supported.

There is enormous scope for innovative enhancements to, and modifications of the interface with this culturally important technology. Really to enable these improvements the development of open and standardised television architectures is required. This would facilitate the development of a secondary market in accessible technology, perhaps through the adoption of solutions already available in the personal computer market, and perhaps of wholly new solutions specialised for television.

2. BRIEF TECHNICAL DESCRIPTION OF THE WORKSHOP

The aim of the workshop is to support inclusive design within the interactive television industry by exploring the development both of methods for its incorporation in the design process and technology for its realisation, which are effective and commercially acceptable.

The immediate goals of the workshop are:

- To collate and summarise the issues that inclusive design raises for digital and interactive television.
- To explore new methods for inclusiveness in the design process.
- To further establish a multi-disciplinary community with the aim of influencing research, development and policy relevant to inclusive and accessible DTV.

3. WHY AND TO WHOM THE WORKSHOP IS OF INTEREST

The workshop will be of interest primarily to designers, broadcasters and researchers concerned both with the facilitation of accessible features in current technology (particularly with regard to the digital switchover) and longer term to the innovative provision of TV-Centric services for inclusion. It will also be of interest to those involved with standards and policy issues for accessibility and inclusion.

There is—unfortunately—a history of divergence between those researchers and designers for whom the topic of inclusive design

naturally holds a vital interest, and those for whom it should. The associated concepts of 'accessibility' and 'disability rights' have an aura of worthy dullness for many designers and business developers; a topic that must no doubt have some attention paid to it (if only to avoid potential legal consequences), but that must not be allowed to get in the way of 'creativity'. This is *so* unfortunate, both for those who want or need to use the products and services, but have not been catered for in their design, *and* for the companies who are missing out on the patronage of an increasingly significant proportion of an ageing population. Also the challenge of including all potential customers in the demographic designed for requires the application of fresh and novel ideas; real creativity that is likely to pay off in increased general usability and acceptability.

The workshop will be of interest to researchers and designers in the digital television industry who are already committed to the cause of inclusive design. In the publicity for the workshop we will also attempted to make the relevance of the issues clear to a wider circle of designers and technologists, by pointing out the significance of inclusive design to current DTV architecture projects such as the BBC Canvas [1] and the Kendra Initiative [2].

This workshop follows that run at EuroITV 2008 in Salzburg and is intended to take forward the ideas developed there. It should be seen as part of the developing agenda for inclusivity to be positively addressed in the design of all software enabled artefacts.

4. PRE-WORKSHOP ACTIVITY

To register for the workshop participants were asked to submit position papers. Digital television industry professionals who have not have previously worked on accessibility or inclusivity were also encouraged to participate; in which case their position paper may raise questions or issues about the application of inclusivity to their particular specialism in the DTV industry.

Suggested themes included but were not limited to:

- Strategies for ensuring accessibility in the digital switchover
- Design for diversity (rather than an 'average user')
- Design for supporting users with specific impairments
- Inclusive and participatory design approaches
- Design for ubiquitous TV
- Tangible interaction and iTV
- New metaphors and models for iTV interaction

The position papers were made available via a web site. This also held links to relevant material; workshop participants were invited to contribute links to draw attention to relevant work and developments.

5. WORKSHOP AGENDA

The half-day workshop will begin with a scene setting presentation on contemporary research into DTV accessibility issues.

Based on the previously submitted position papers a set of challenges for discussion will be prepared and circulated to the participants in the weeks before the workshop. These will be used to initiate discussion, with the emphasis being on practical

outcomes that may be pursued. The final challenge to the workshop will be to address the practicalities of putting a research agenda into action. The progress of initiatives for collaboration through FP7 and other funding streams developed during the workshop at EuroITV08 will be reported.

6. POST-WORKSHOP ACTIVITY

Notes on the discussion will be placed on the workshop web site, which will remain accessible on the web indefinitely [3]. Participants will be encouraged to contribute reflective commentaries.

It is intended that practical outcomes in the shape of research proposals and potentially, initiatives aimed at standards, industry and regulatory bodies will be developed out of contacts made at this and the previous EuroITV 2008 inclusivity workshop.

7. WORKSHOP ORGANIZING COMMITTEE

Mr Richard N Griffiths a member of Interact (the British HCI Group) is course leader of the MSc in Digital Television Management & Production at the University of Brighton, UK, a course the he has managed since its inception in 2001. He researches usability design for interactive TV, particularly accessibility, and is currently involved with an EU FP6 multi-partner project; LOGOS: Knowledge on demand for ubiquitous learning. He has previously been an organising member of workshops at Interact 1999 and CHI 2000 on usability pattern language, and HCI 2002 and 2003, and EuroITV08 on iTV and accessibility.

Dr Mark Springett is a member of the Interaction Design Centre and the Design-for-all research group at Middlesex University. He is a member of the Usability Professionals Association. He is a Working Group Co-ordinator for MAUSE (COST Action 294). He has 20 years experience of working in the area of Human-Computer Interaction, and has had a specialist interest in accessibility since 2000. His recent research includes investigations of iTV accessibility for citizens with low vision. He Co-chaired the workshop 'Inclusive Interactive Television: Setting the Agenda for Innovative Research' at EUROITV 2008. He was Co-organizer of the COST 294 workshop 'Valid Useful Usable Evaluation Measures' in Iceland in June 2008.

7.1. Primary contact

Dr Mark Springett

8. REFERENCES

- [1] Informitv, 2008. BBC opens Canvas for platform-neutral publishing. DOI=<http://informitv.com/articles/2008/10/14/bbcopenscanvas/>
- [2] The Kendra Initiative. DOI= <http://www.kendra.org.uk/>
- [3] EuroITV 2009 Workshop: Defining the Architecture for Next Generation Inclusive Television http://www.cmis.brighton.ac.uk/staff/rng/EuroITV/09_Inclusivity_Workshop/

Inclusive T-gov Application Development in Brazil

Lara Schibelsky Godoy Piccolo, Daniel Moutinho Pataca
CPqD – Telecommunications Research and Development Center
Rod. Campinas Mogi-Mirim, km 118,5
Campinas, SP – Brazil
+55 19 3705-5809

{lpiccolo, pataca}@cpqd.com.br

ABSTRACT

This paper is part of a research project that aims at creating an interactive digital TV multiplatform service, bringing e-gov inclusive and accessible applications developed for Web to TV. The defined target audience is the Brazilian society, which includes low literacy people and people with disabilities. The universal design approach is taken into account when proposing solutions that considers usability, intelligibility and accessibility concepts are equally important to reach an inclusive t-gov service. Some approaches to develop these applications are presented.

Keywords

Accessibility, T-gov, universal design.

1. INTRODUCTION

Different from most European countries, the free-to-air digital TV was recently launched in Brazil, and pay TV, that traditionally has a low penetration level, has been popularized. These facts reinforce that interactive TV is still a promising tool to help bridge the digital divide [1].

To reach this ambitious target, it is essential to consider the inclusivity of this new technology. This concept is related with the possibility for accessible and enjoyable technology for **everyone**, beyond the commitment to improve the quality of life for the elderly as well as to people with disability [2][3].

Although this moment of consolidation of the technology can be considered the perfect time for proposing technical solutions which consider the Brazilian population's needs within its widest extension, all the digital TV industry efforts in the short run are concentrated to expand the high definition video transmission, and to make the interactivity feasible itself.

Beyond these efforts, this position paper is part of continuing research in Brazil that proposes to create a set of interactive digital TV (iDTV) multiplatform services bringing e-gov inclusive and accessible applications developed for Web to TV. Therefore, the project's background and its intent will be presented, as well as some possible solutions to deal with the technological constraints regarding accessibility imposed by the moment.

2. BACKGROUND

Free-to-air represents more than 90% of the TV sets in Brazilian households. Presently, the digital transmission is available in 10 main cities, broadcasting high definition video content. The population cannot enjoy the interactivity until the Brazilian middleware (GINGA) has been integrated to set-top box by manufacturers and the value chain to offer interactivity has been defined. Although digital cable and digital satellite TV are becoming more popular, pay-TV still has low penetration (around

10%). Telecom operators have recently been allowed to offer TV services as well, creating new offers to the population.

In previous research [3] [4], an analysis of the iDTV accessibility in the Brazilian context was taken. It considered informal, formal, and technical aspects and resulted in a set of recommendations for design accessible interfaces by referring to the W3C guidelines 2.0 for Web and specific iDTV recommendations.

According to the analysis, the current regulations are not enough to ensure that iDTV contents will be accessible to the population as a whole. Representative groups of people with disabilities still require the right to access part of broadcasted analogue TV content with good quality assistive resources (closed caption, audio description, and sign language window). The impact of an interactive contents offer with more textual information, and new interaction models, is still unknown and unmeasured by these groups that will only take it into account when the technology is ready to consume.

Among other issues, the SMTVI¹ project may subsidize these groups of users with information about what they can expect and require from this technology; and may contribute with the iDTV industry proposing possible solutions and recommendations about how to design and to offer inclusive iDTV applications, considering the impacts to the TV receiver hardware and middleware.

2.1 The project purposes

The SMTVI project aims at developing services that encourage people without access to other information and communication technologies to use it.

According to [5], 47% of the Brazilian population has never used a computer and 84% of those people are illiterate. In 2005, nearly 74% of the population had low to medium literacy skills [6]. Sensory, motor or physical disabilities afflicted 14.5% in 2000 [7] [8]. The project stands to conceive inclusive iDTV multiplatform services, such as games, t-learning, t-commerce, and t-gov services, according to the universal design approach. The user interfaces consider usability and intelligibility issues, intending to be flexible enough to be fully enjoyable by the entire population, including people with sensorial disabilities; beyond that, it must be attractive and easy to use, even to first-time users.

3. INCLUSIVE T-GOV APPLICATIONS

To specify a t-gov application it needs to consider the population at large as the target-audience. An experience indicated that literacy levels seem to influence the experimentation and

¹ SMTVI (Interactive Digital TV Multiplatform Services) is founded by the Brazilian Communications Ministry.

appropriation of the technology by the user more than the user's age, but the collective use may help to alleviate this barrier [10].

To preserve the collective way to enjoy TV, the universal design approach was considered adequate for this. Thus, people with different abilities, whether resulting from aging, illiteracy and disability or not, will benefit from accessible products and services, which do not discriminate against them [3].

Usability and intelligibility concerns to this target-audience were previously studied in Web domains according to the user-centered design approach [8], resulting in the creation of proper metaphors and linguist aspects applied to an interaction model (see Fig. 1). To bring the interaction model close to the user experience, many aspects of TV communication language was applied.



Fig. 1 - Screenshot of the e-gov service for Web

This knowledge will be adapted to iTV specificities in a service called GTV, composed by three t-gov applications: a health service for making doctors appointments and to offer health-related topics contents; an information service about social security; and a job portal that allows the user to send and receive messages from the employers. This service can be offered by IPTV, cable, satellite or broadcast, as illustrated on Fig. 2 [9].

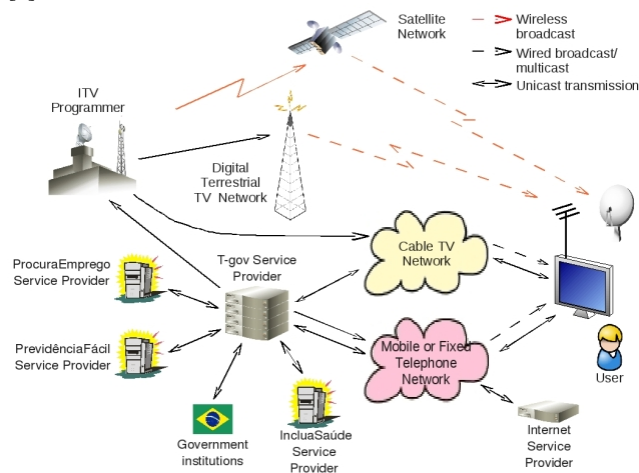


Fig. 1 - GTV architecture

4. THE ACCESSIBILITY APPROACH

It is currently not possible to count on assistive resources such as embedded text-to-speech converters or automatic generation sign language windows on the Brazilian set-top boxes. Intending to follow the recommendations [3] [4], the SMTVI approaches to manage these constraints are:

- To offer pre-recorded audio to all interactive options;
- To produce inclusive audiovisual content with captions and sign language



Fig. 2 - Inclusive content

of equal size on screen (see Fig. 3);

- To make audiovisual the primary tool above textual content;
- To evaluate an automatic sign language generation with avatar on the server side of the service;
- To evaluate the execution of a text-to-speech solution on a pen-drive plugged into the set-top box.

These possible solutions must be implemented and evaluated with potential users. The last two possible solutions must be evaluated both technically and with potential users.

5. CONCLUSION

Usability, intelligibility and accessibility are equally important to reach an inclusive technology. Taking into account the technical and regulatory constraints, some approaches to conceive a set of accessible and inclusive t-gov applications for the diversity of the Brazilian society have been presented. After development, the next step is to evaluate the results with potential users, remembering that it may be necessary to create new methods of evaluation that focus on the intelligibility and affective relation of the user with the technology.

6. REFERENCES

- [1] Kutiishi, S. M., Piccolo, L. S. G. 2006. Interactive TV Services and Applications for Digital Inclusion in Brazil. In Proceedings of the EuroITV 2006.
- [2] Connell, B. R., Jones, M. *et al.* 1997. About UD: Universal Design Principles. Version 2.0. Raleigh: The Center for Universal Design.
- [3] Piccolo, L. S. G., Melo, M. A., Baranauskas, M. C. C. 2007. Accessibility and Interactive TV: Design Recommendations for the Brazilian Scenario. In Human Computer Interaction - INTERACT 2007. 11th IFIP TC 13 International Conference, Rio de Janeiro. Lecture Notes in Computer Science 4662. Berlin: Springer, 2007. v1. 361-374.
- [4] Piccolo, L. S. G., Melo, A. M., Baranauskas, M. C. C. 2007. A Convergent Proposal for Accessible Interactive TV Applications Development. In Interactive TV: A Shared Experience TICSP Adjunct Proceedings of EuroITV 2007.
- [5] CGIBr, 2007. Pesquisa sobre o uso das tecnologias da informação e da comunicação no Brasil. <http://www.cgi.br>.
- [6] Instituto Paulo Montenegro www.ipm.org.br.
- [7] Brazilian Geography and Statistics Institute – IBGE www.ibge.gov.br.
- [8] Ávila, I., Menezes, E. *et al.* 2007. Interaction Models for Digital Inclusion of Low literacy, Aged and Impaired Users in Brazil. In Proceedings of the 3rd Latin American Conference on Human-Computer Interaction - CLIHC 2007.
- [9] Pataca, M. D., Piccolo, L. S. G. 2009. Especificação Funcional. Serviço T-gov. PD.30.12.34A.0010A/RT-01-AA. CPqD. FUNTEL.
- [10] Piccolo, L.S.G., Baranauskas, M.C.C., Furtado, M.E.S. 2008. A T-gov Prototype Evaluation Considering a Wide Diversity of Users. Proceedings of the IADIS Interfaces and Human Computer Interaction - IHCI 2008.

Personalization for digitally excluded users through t-learning

Pablo Rebaque

Barcelona Media – Center of Innovation
Av. Diagonal 177, 9th floor, 08018 Barcelona
+34 932381400

pablo.rebaque@barcelonamedia.org

Raquel Navarro-Prieto

Barcelona Media – Center of Innovation
Av. Diagonal 177, 9th floor, 08018 Barcelona
+34 932381400

raquel.navarro@barcelonamedia.org

1. POSITION PAPER

In recent years several researchers have provided evidence about the impact of personalization on TV [1, 2, 3, 4]. Results show the consumer profits when the content fits a given user profile, age range or social group (independently from viewing device - Internet, TV, mobile, etc). The benefit is related to a higher level of acceptance of content, higher enjoyment, and interest, that is, user engagement.

Also, the application of ITV for learning (t-learning) has emerged with growing interest. For instance, in Spain several studies focused on t-learning have been conducted, dealing with technological platform development and more specifically the development of an ontology that allows linking educational content to entertainment, where this content would be personalized. However, the parameters that have been taken into account in order to define the personalization are not well known. Furthermore, it is unknown if other variables have been considered to provide personalized content.

The discussion topic we would like to bring to the workshop is how to integrate the possibilities of personalization through ITV, educational content through ITV, and a new way of interacting (due to the fact that the content would be presented through TV) into a unique model.

More specifically, our contribution to this workshop aims at the less technical user, such as the elderly and those who fall into the digital divide (+55 in Spain). Considering that this type of user has a large experiences on TV as compared to a small expertise over other technologies such as the Internet, it would seem interesting to focus on how this public perceive the possibility of using a technology such as t-learning to acquire new knowledge. For instance, how to offer the content personalization to this user group, or what type of knowledge they would be interested in purchasing, or whether they would be entertained with the same content items as younger users, and so on.

This workshop could generate discussion about a number of issues or challenges according to different usage scenarios that we are investigating for digitally excluded people:

IPTV usage scenario

How can users profit from the elements from Web 2.0. (for instance, labelling group) in ITV? How can we gather such

information? How could we present this information to users? How to give feedback?

Study the impact of avatars as a mechanism for interaction or as a mechanism for presenting information. How could we adjust personalization and socioemotional effects (to achieve a high level of engagement and entertainment) depending on the avatar?

Analyze the adequacy of formats for a satisfying user experience. How are the contents presented? How to interact with the content? Which is the best way to learn?

The InternetTV usage scenario

Definition of a successful experience for the digitally excluded user in terms of:

- The adequacy of the Web 2.0 possibilities for the large amount of content available on the Internet
- The multimedia complementarity in terms of:
 - the content topic
 - format (video, image, text ...)
 - user goals
 - personal variables
 - the user context (time, place, alone or accompanied)

Study the impact of avatars as a mechanism for interaction or as a mechanism for presenting information. How could we adjust personalization and socioemotional effects (to achieve a high level of engagement and entertainment) depending on the avatar?

Considering the possibilities of Web 2.0, the large amount of content available on Internet, and aiming at users having a successful experience: what should the formats for content presentation be? How should the user-content interaction be?

Usage scenario of convergence of IPTV and InternetTV

In what ways do the Internet (lean forward) and TV (lean back) converge? What elements do they have in common and what they do not? - Is it possible to create continuous forward momentum for the user experience?

2. REFERENCES

- [1] Bonnici, S. 2003. Which channel is that on? A design model for electronic programme guides. In Masthof, J., Griffiths, R., and Pemberton, L., editors, *Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors?* Brighton, UK, pages 49–57. (Available online: <http://www.itu.dk/courses/PDI/F2005/Paper6.pdf>)
- [2] Chorianopoulos, K. et al, 2003. The Virtual Channel Model for Personalized Television. *Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors?* Brighton, UK, pp.59-67. (Available online: <http://www.brighton.ac.uk/interactive/euroIPTV/euroIPTV03/Papers/Paper7.pdf>)
- [3] Sperring, S. & Strandvall, T., 2006. exEnteractivity – Experiencing Entertainment in Interactive Television. In: Paper presented at the *NordForsk doctoral course “Publics, audiences and users”*. Køge, Denmark. (Available online: http://www.ruc.dk/upload/application/pdf/f51d6748/Susanne_Sperring_and_Tommy_Strandvall_Paper.pdf)
- [4] Van Dijk, J. et al, 2003. Interactive television or enhanced television? The Dutch users interest in applications of IPTV via set-top boxes. In: *Paper Presented at the 2003 Annual Conference of the International Communication Association*. San Diego, USA. (Available online: http://www.gw.utwente.nl/vandijk/research/IPTV/IPTV_plaatje/a_interactive_of_enhanced_tv_i.pdf)

VestibaTV – An Interactive Program for Vestibular Training

André Felipe Palmeira	Marcelo Fernandes de Sousa	Tatiana Aires Tavares
Digital Video Applications Lab (LAVID)	Digital Video Applications Lab (LAVID)	Digital Video Applications Lab (LAVID)
Federal University of Paraiba	Federal University of Paraiba	Federal University of Paraiba
Computer Science Department	Computer Science Department	Information Science Department
+55 83 3216 7093	+55 83 3216 7093	+55 83 3216 7093

andre@lavid.ufpb.br

marcelo@lavid.ufpb.br

tatiana@lavid.ufpb.br

Ricardo Mendes C. Segundo
Digital Video Applications Lab (LAVID)

Federal University of Paraiba
Computer Science Department
+55 83 3216 7093

ricardo@lavid.ufpb.br

Guido Lemos de Souza Filho
Digital Video Applications Lab (LAVID)

Federal University of Paraiba
Computer Science Department
+55 83 3216 7093

guido@lavid.ufpb.br

ABSTRACT

In this paper, we discuss our experience of developing an interactive TV program for training people to vestibular. Many social organizations consider the vestibular model a barrier to public further education access in Brazil. One way to better the participation of people from underprivileged classes can be creating and offering preparatory t-learning courses taking advantage of TV range in Brazil.

Categories and Subject Descriptors

D.4.7 [Organization and Design]: Interactive systems

General Terms

Design, Experimentation, Human Factors.

Keywords

Interactivity, Digital TV, Social Aspects of Software.

1. INTRODUCTION

Digital technologies have opened new directions for experimentation in the field of Learning – in order to facilitate its. The use of digital technologies can reach social groups which have no access to traditional forms of Education. Using new mechanisms of Education we can go where people are, for example, using a TV we can go inside people homes (Collete, 2001) [1]. In Brazil, according to *Eletrobras* research, 97.1% of the households have a television, but, according to IBGE, only 20% of them have computers with Internet access [4]. So, one of these technologies that has a powerful mode to reach people is the **Interactive TV** (iTV), which combines Internet and TV through the *t-learning* applications.

The term t-learning comes from the convergence between iTV and Internet e-learning techniques. But this new technology has its own specific characteristics, for instance, a low screen resolution, the use of a simple remote control to interact with programs and the decreased number of features in a set-top box compared to a computer [5]. Another motivation for t-learning is the problematic of the access to universities in Brazil by the

poorest social classes. Researches show that in Sao Paulo, for example, 85% of students are in the secondary school, but only 20% of them are approved by the FUVest vestibular (University Foundation for Vestibular) [2]. Many social organizations consider the vestibular model a barrier to public further education access for these classes. Besides, there are specialized schools that train students to pass the vestibular examination, using classes and simulation tests. But this kind of school is very expensive and mostly accessible by the higher classes of the population. Other way to better the participation in the vestibular examinations can be creating and offering preparatory t-learning courses which can be accessible for anyone who has a TV set at home (the majority of Brazilian population), consequently enabling digital and social inclusion.

In this paper we present a t-learning application for vestibular training that we called VestibaTV. VestibaTV was developed in the context of BEACON project and it aims to achieve some features as: (a) using an accessible language to juvenile audience; (b) presenting interactivity contents; (c) improving simulation exercises and, (d) developing topics of awareness and citizenship for the students and the community at all [3][5]. In addition, we are proposing an architecture for cross media content based on VestibaTV results.

2. VestibaTV ITV PROGRAM

In this section we briefly describe the main models developed to VestibaTV. First we present the user needs, the main functionalities and the user interface for the ITV program. Besides, we discuss the proposed architecture for VestibaTV cross media content.

2.1 User Needs and Functionalities

Figure 1 shows the use case model for VestibaTV. Note that the use cases for the *student* actor were implemented by the interactive TV program. The *teacher* use cases are being included to the cross media approach. Regarding that we are working with users from underprivileged social classes and the simplicity is one of the most important features of the application. We have to offer something easy to use that can attract the attention of the user without disincenive them.

The main offered features are: (1) **Video content**: the video content has to use a simple language and to show situations of daily life where the student can see their own reality; (2) **Tips**: while the video is exhibited an icon indicating the possibility of interaction appears several times. When clicked, a tip with information about the topic addressed in the video appears in textual to fix the concepts presented during the video; (3) **Vestibular Simulated**: when the video ends, the viewer is invited to start a vestibular simulated. This is a review mechanism where vestibular questions and four answer alternatives are showed to viewers. If the viewer chooses the correct alternative, additional information regarding the issue will be displayed along with the congratulations. Otherwise, the correct answer is presented to the viewer.

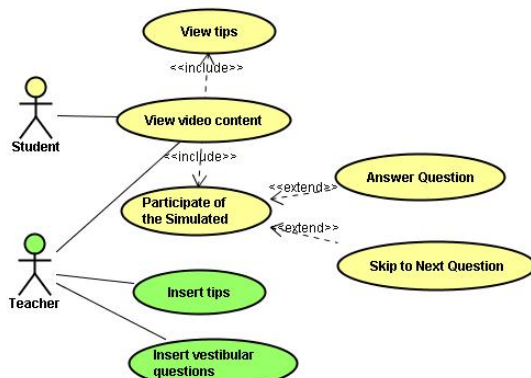


Figure 1 – Use Case Diagram for VestibaTV.

2.2 User Interface

VestibaTV interface is very friendly and intuitive. All the interactivity is navigable by the color buttons (red, green, blue and yellow) of the TV remote control. For guiding the users during the interaction we use a friendly owl figure.

For example, Figure 2 shows the user interface of the VestibaTV simulated. Note the presence of the application guide (the little owl) and the color buttons for interacting with the application. VestibaTV was developed in Ginga-NCL and it is running in a Proview receptor with Ginga-Ready.

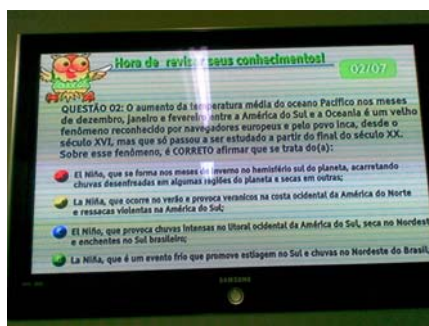


Figure 2 – VestibaTV Simulated User Interface

2.3 Proposal Architecture

Figure 3 shows the proposed architecture for interactive applications involving the Web Server, the TV Station, the PC

and the Set-Top Box. The communication between the Web Server and the PC occurs by a bidirectional form, through uploads (including updates) and downloads. We have the teacher profile where teacher could submit new content to the server. At the other hand, we have the user profile where students are allowed to download the video content and the applications (Web Server) to their home PCs. Besides, students can download applications on removable disks (Pendrive) and execute them in the STB USB input. The applications can be downloaded directly from the IP network connection without user mediation. Finally, we have communication between the TV Station and the STB, which is made through a diffusion channel.

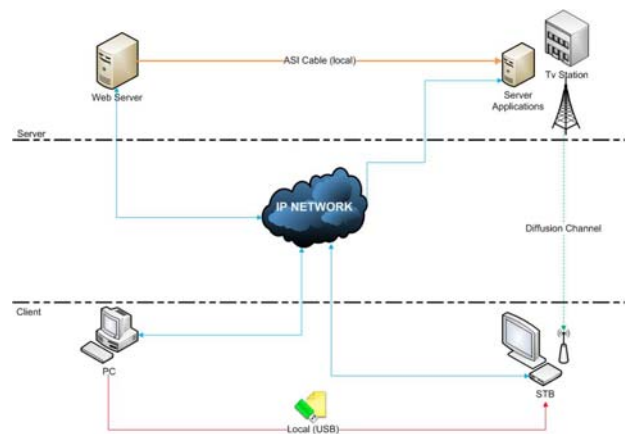


Figure 3 – Communication Diagram

3. CONCLUSIONS

In this article, we presented a t-learning application for vestibular training in Brazil, as well as, the proposal architecture to a cross media model for Digital TV and Web applications. We believe that the t-learning can be the suitable medium to foster an inclusive society by supporting strategies to combat social exclusion, inequalities and the discrimination of different representations of social values.

4. ACKNOWLEDGMENTS

Thanks to European Commission for the Beacon project support.

5. REFERENCES

- [1] Lytras, M., Lougos, C., Chozos, P., and Pouloudi, A., 2002. Interactive Television and e-learning convergence: examining the potential of t-learning. Available online: <http://www.eltrun.aueb.gr/papers/tlearning.pdf>.
- [2] MITRULIS, E.; PENIN, S. T. S. Pré-vestibulares alternativos: da igualdade à equidade. Cadernos de Pesquisa, São Paulo, v. 36, n. 128, p. 269-298, maio/ago. 2006.
- [3] BEACON – Brazilian-European Consortium for DTT Services. <http://www.beacon-dtt.com/en/index.php>.
- [4] Eletrobrás/Procel - <http://www.eletrobras.gov.br>
- [5] Rey-Lopez, M., Diaz-Redondo, R., Fernandez-Vilas, A., and Pazos-Arias, J. 2007. Entercation: engaging viewers in education through TV. ACM Entertaint. Vol. 5, No. 2, Article 7.
- [6] BEACON UFPB – www.lavid.ufpb.br/beacon

Interactive Digital Television to enhance older people's quality of life

Andrea Miotto

Psychology Department,
Goldsmiths, University of London
New Cross, London, SE14 6NW, UK
(+44/0) 20 7919 7338
A.Miotto@gold.ac.uk

Jane Lessiter

Psychology Department,
Goldsmiths, University of London
New Cross, London, SE14 6NW, UK
(+44/0) 20 7919 7338
J.Lessiter@gold.ac.uk

Jonathan Freeman

Psychology Department,
Goldsmiths, University of London
New Cross, London, SE14 6NW, UK
(+44/0) 20 7919 7338
J.Freeman@gold.ac.uk

ABSTRACT

As people grow older, their cognitive ability tends to decline, which can result in mental, physical and social hardships. Recent research has shown that cognitive training can provide a cognitive stimulation which slows down the decrement in capacities related to the ageing process. Our work aims at using cognitive psychology and Interactive Digital Television applications to enable older people to actively and autonomously participate in brain training activities while sitting in front of their TV set. In this paper, we present an initial overview of the user requirements useful for the technical design stages of the interactive system and key factors that may facilitate or hinder the adoption by older people of brain training applications.

Categories and Subject Descriptors

A1 [Introductory and Survey]

General Terms

Design, Experimentation, Human Factors, Measurement.

Keywords

Human Computer Interaction, usability, user needs, use cases, user-centered design, drivers and barriers, television, TV, mind fitness, brain training, cognitive training program, inclusion, elderly, older people.

1. INTRODUCTION

Population ageing is increasingly being recognised as one of the most salient social, economic and demographic phenomena of the near future [1]. It is estimated that by 2050, one-third of Europe's population will be over 60, compared to 13 percent who will be under 16 [2].

While the significant growth in the numbers of older people in nearly every country in Europe has been widely publicised, it is not sufficiently recognised that this growth also implies significant increases in the proportions of older people who remain healthy and active well into their seventies, eighties and beyond. In general, it is important to note that even healthy old age is accompanied by an accumulation of mild but progressive losses of efficiency in all sense organs [3].

In general, the older a person is, the more likely he or she is to suffer impairment in cognitive functions. According to Alzheimer's Disease International [4], in people younger than 65 years of age, dementia affects about 1 person in 1000. In people over the age of 65, the rate is about 1 in 20, and about 1 in 5 people over age 80 have dementia.

2. COGNITIVE TRAINING TO REDUCE AGE-RELATED DECLINE

Whilst cognitive decline becomes increasingly more common with advancing age, a growing body of research has investigated whether cognitive training may delay the onset of or even prevent the changes associated with such decline [5, 6].

The most famous study in this field was conducted by Ball [7] with a sample of 2832 participants aged 65 to 94 years. The results of this study indicated that computerised cognitive training programmes applied to older people resulted in long-term improvements in targeted cognitive abilities.

Although previous studies have shown that using brain training on a PC can help improve the quality of life as aging occurs, the majority of older adults is not computer literate and cannot benefit from the growing number of mind fitness training programs designed to preserve cognitive ability. Instead, older people resort to more passive ways of spending leisure time, such as watching TV, which tend to be risk factors rather than protective ones to cognitive ability [8].

3. RESEARCH PURPOSE

The research presented in this paper aims to investigate the use of computerized cognitive training via interactive digital television (iDTV) to ameliorate the declines associated with later life.

The purpose of this paper is to give an overview of the initial findings of research into the user needs that should be considered in order to specify requirements in the area of cognitive training for older people and the potential drivers and barriers to the adoption of brain training applications via iDTV. Inspiration is drawn from both existing tools used for cognitive training in computer environments and the literature on Human-Computer Interaction to inform specific applications for iDTV.

4. METHODOLOGY

We started investigating these objectives in two ways. First, we developed user scenarios with personas to brainstorm on what we envisage a brain training application should provide. Second, we conducted a series of preliminary focus groups with older people to collect information about the needs of older people and to investigate what can motivate or hinder them to adopt and use iDTV applications that help to improve some cognitive functions.

5. INITIAL FINDINGS

Based on the result of this research work, we learned that cognitive training delivered via iDTV may appeal to older people, especially for those who have little experience with computers and who lead a less active lifestyle. Many people of retirement age who face the challenge of what to do with their increased free time would be happy to use a friendly device such as the TV to keep their minds active. However, some oldest-old in our study (people aged 80 years and over) were more likely than their younger counterparts to be easily intimidated by new technology, especially if they have little or no experience of technology, and they may have difficulty in accepting new products that are difficult to use and understand.

Focus groups participants reported that they would like these applications to be used as means of social cohesion, to share activities with family members and friends, while providing privacy to individuals. In fact, whilst some older people reported a particular interest in maintaining close existing contacts with relatives and friends, others expressed their doubts about the idea of communicating with people unfamiliar to them.

The result of the focus groups also indicated that older people need to be constantly motivated and reminded of their goals. Besides the use of motivational feedback, additional information are necessary to support and coordinate their activity and assist the user in interacting with the system. Finally, while some older people like the idea of receiving feedback on their task performance, others made it clear that they would rather not know the extent to which their cognitive abilities were declining.

6. FUTURE WORK

A pen and paper questionnaire will be used to survey a large sample of older people. We plan to structure this questionnaire in several sections:

- An inventory of 40-50 attitudinal questions, designed to explore the attitudes of older people toward leisure activities and technology and to develop a typology of potential users of brain training applications.
- A series of 10-15 demographics questions, which will be useful along with the attitudinal profiles to create a rich picture of the potential user groups.
- Finally, a series of questions on the use and access of technology to test and further refine the key factors that may facilitate and prevent the adoption of such applications.

7. ACKNOWLEDGMENTS

This position paper is part of a project called "Vital Mind", a research programme funded by the European Commission under the Seventh framework Programme for Research and Technological Development and its specific programme 'Ambient Assisted Living'.

8. REFERENCES

- [1] Eurobarometer. 2005. Special Eurobarometer 223: Social Capital. European Commission.
- [2] Špidla, V. 2005. Europe's population is getting older. How will this affect us and what should we do about it?. Press release IP/05/322 on the 17th of March 2005.
- [3] Carmichael, A. 1999. Style guide for the design of interactive television services for elderly viewers. Independent Television Commission, Winchester.
- [4] Alzheimer's Disease International, 1999. The Prevalence of Dementia (Factsheet). London.
- [5] Gunther, V. K., Schafer, P., Holzner, B. J., & Kemmler, G. W. 2003. Long-Term Improvements in Cognitive Performance through Computer-Assisted Cognitive Training: a Pilot Study in a Residential Home for Older People. *Aging & Mental Health*, 7, 200-206.
- [6] Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., et al. 2006. Memory Enhancement in Healthy Older Adults Using a Brain Plasticity-Based Training Program: A Randomized, Controlled Study. *The Proceedings of the National Academy of Sciences*, 10.1073/pnas.0605194103 (Medical Sciences).
- [7] Ball, K., Berch, D., Helmers, K., et al., 2002. Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of American Medical Association*, 288(18):2271-2281.
- [8] Sanders, A.E., Zimmerman, M.E., Verghese, J. 2007. Cognitive correlates of television programming preferences. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 3(3):1.

Digital Television and the Open Mediaspace

Kari-Hans Kommonen
Media Lab, University of Art and Design Helsinki
Hämeentie 135 C
00560 HELSINKI, Finland
+358 9 7563 0563
khk@taik.fi

Sanna Marttila
Media Lab, University of Art and Design Helsinki
Hämeentie 135 C
00560 HELSINKI, Finland
+358 9 7563 0412
sanna.marttila@taik.fi

POSITION PAPER

The digitalization process in society has created a tremendous potential for the evolution of new forms of media, by turning the media environment that used to be formed of rigid, strictly controlled channels, into a flexible, open Mediaspace.

The Mediaspace is based on the computer, a metamedium that can be programmed to support any type of media, and on the Internet, the global communications network that is open and being made accessible to everyone. The computers take many forms, such as desktop and laptop computers, PDAs, mobile phones, set-top-boxes and digital television sets. The Internet extends to these computers also in many forms, through fixed telephone cables that have been taken up by ADSL, dedicated Internet cables, WLAN, wireless GPSR or UMTS signals, or digital broadcast transmissions, such as DVB and DVB-H.

However, in this flexible Mediaspace, the digital television has remained as a system of rigid channels and fixed functionality. The digital television system – its hardware, networks, economics, production and distribution systems – has been designed by the actors in the television business who have quite naturally seen the new digital system as a continuation of their existing business and activities, and have overlooked or knowingly dismissed the potential for new arrangements in the social production and use of media that the new technologies and distribution topologies could have offered, if configured differently.

The world around the digital television system, however, has evolved to embrace new interesting possibilities. Digital production tools, extending from built in webcams in laptops, digital still and video cameras in mobile phones to HDTV video cameras costing a few hundred Euros, combined with cheap editing software in home computers have made individual video production a real and massive source of audiovisual media in society.

Countless YouTube-like services and peer-to-peer networks have created an efficient and free infrastructure for the distribution of any type of media globally. As these services exist in the open ecosystem and often offer easy-to-use construction interfaces that let other services embed their offerings in new formats, they have become building blocks for new media applications.

Social media that combines media production with social enrichment features that allow communities of people to select, recommend and comment media, has co-evolved with the massive diversity of content production and made it possible for the "long tail" of user created media to find its communities of interest, and as these communities emerge, they further motivate their members to engage in media production and social processing that is meaningful for them.

This is the most interesting area of media evolution at the moment, but unfortunately, because of its rigid and controlled design, the digital television has been almost completely left out of this development and confined in its existing configurations.

While the DTV industry has shown little signs of interest to change the digital television system, it appears that the open world may evolve a way around the rigid system, as there is an increasing repertoire of hardware and software solutions for replacing the limited functionality set-top-boxes with open, programmable computers that open up all of the possible media channels and distribution and selection systems to television audiences. Examples of these include the AppleTV, Elisa, Boxee or XBMC.

The large commercial content owners have recognized the potential of Internet distribution of their content, and hence audiences are getting new ways to access also their offerings through the open system. Eventually, this development may lead to a situation where the open replaces the closed system.

For the development of the future digital television, this is a real challenge. The power of the system designers to control the way how their system is embedded in the everyday life and the media environment is disappearing. It appears that it would be a better strategy if the next generation of digital television would be designed to take advantage of the potentials of the Mediaspace.

In terms of interaction, inclusion and accessibility, embracing an open architecture and designing for an open digital television ecosystem will also open up the possibility of open design for custom interfaces and interaction models for and by specific communities. A more open television environment will have the potential to cater for specific communities also in terms of both content and functionality.

While we are not specifically addressing the inclusive access and interaction features of digital television, we propose that it is necessary to see that digital television as it is today, is seriously constrained as a media environment to passive consumption of media, and that upgrading the television to become a universal platform for open media interaction – viewing, navigating, selecting, recommending, organizing, commenting and producing – is a central concern for all television user communities.

In our contribution to the workshop, we would like to

- present examples of what kinds of new media practices the next generation of digital television should facilitate
- suggest design and architectural requirements for the next generation of DTV to make this possible
- discuss these issues with the other participants.

URC Based Architecture for Next Generation Inclusive Television

Gorka Epelde
VICOMTech – IK4
Mikeletegi, 57
20009 San Sebastian, Spain
+34 943 309230
gepelde@vicomtech.org

1. POSITION PAPER

This paper presents a new architecture to make today's and next generation's Television accessible for all. Our proposal is based on the ISO/IEC 24752 "Universal Remote Console Framework" standard [1]. This standard defines an abstract user interface layer called the "user interface socket" and allows the development of pluggable user interfaces for any type of user. Since this architecture is standards based, the development of compatible pluggable interfaces is open to any third party.

Besides, the Universal Control Hub (UCH) is a gateway oriented architecture for implementing the Universal Remote Console (URC) framework in the digital home [2].

The main features of the UCH are:

- It acts as a bridge between targets and controllers: each with its own communication and control protocol, that otherwise would be unable to talk to one another.
- Standard-based user interface socket: The UCH is based on the URC framework previously described.
- A variety of user interface protocols: The UCH allows different user interface protocols (DHTML over HTTP, Flash, etc.) to be implemented and used by controllers.
- Globally available resource servers: The UCH can get distributed resources, such as target adaptors, target discovery modules and user interfaces from resource servers.

Figure 1 shows the UCH architecture for the URC standard.

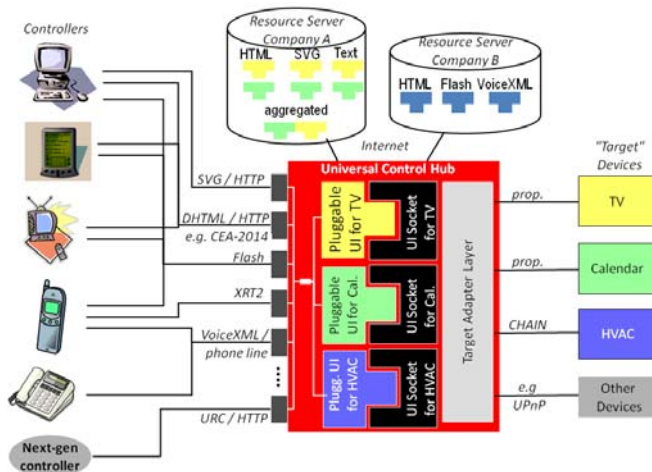


Figure 1. UCH architecture

For more implementation details, have a look to the short paper titled "URC Based Accessible TV" which would be presented at the EuroITV 09 [3].

In two European projects, we have developed and tested this approach to ensure the accessibility of the TV and the digital home. The scope of the first project, i2home [4] is the Intuitive Interaction for Everyone with Home Appliances based on Industry Standards. In this way i2home is making devices and appliances at home more accessible to persons with mild cognitive disabilities and older persons.

The second project, Vital [5] proposes a combination of advanced information and communication technologies that uses a familiar device like the TV as the main vehicle for the delivery of services to elderly users in home environments. The new services offer will depart from traditional assistance schemes, but it considers other important demands, such as: the need for information, the need for inter-personal communication, the need for personal advice, the need for edutainment, the need to be able to move safely in the physical environment and the need to integrate into the mainstream society.

In the i2home project we have made accessible the control of different devices: TV, Calendar system, Siemens home appliances, etc. In the case of TV, two different implementations have been done on the first iteration of the project, the Dreambox 7025 [6] was integrated to the UCH architecture, while in the second iteration Microsoft Vista Media Center [7] was integrated to the UCH.

In this project, several advanced user interfaces have been developed to control the TV and the other devices and services. For example, for the control of the TV, a multimodal user interface running in a PDA adapted to people with cognitive impairments and a accessible DHTML page for the people with sight impairments have been developed.

One of the advanced interfaces deployed on the TV has been the virtual character or the avatar, which has shown good results on tests conducted with persons with Alzheimer's Disease ranging from mild to moderate [8, 9]. An example of this advanced user interface is shown in Figure 2.

In contrast to i2home, the Vital project is centred in providing accessibility to a concrete user group, which is the elderly people without any special impairment, just the ones associated to the aging. In the Vital project, the client devices have been restricted to the TV at home and the mobile phone outdoors. The Vital project concentrates its effort in providing more services to a concrete user group, in contrast to i2home, which follows an "interfaces for everyone" approach.



Figure 2. Virtual character's user interface displayed on TV.

Following the previously stated objectives, the Vital project provides the following services interacted through the TV:

- Videoconference service.
- Entertainment services. For example: Chess game, quiz game,...
- Edutainment services.
- Personalized information services. For example: Personalized news, personalized web navigation,...
- ...

Other pluggable user interfaces for different client devices and users could easily be developed for the services integrated in the Vital project.

In the workshop we will explain our approach in detail. From our point of view, there is no user interface that fits all user groups' needs or preferences. User interaction modalities that are useful for some users could be disturbing or unnecessary for others. So, we believe that the solution consists in providing a standardized and open architecture that allows third parties to rapidly develop pluggable user interfaces adapted to the needs or preferences of each user group. In this way, an open market for these pluggable user

interfaces could be established and people would be able to share or buy different user interfaces. An example of this open market is dotUI [10].

2. REFERENCES

- [1] ISO/IEC. 2008. Information Technology - User Interfaces - Universal Remote Console (5 parts). ISO/IEC 24752.
- [2] Zimmermann, G., and Vanderheiden, G. 2007. The Universal Control Hub: An Open Platform for Remote User Interfaces in the Digital Home. In *Proceedings of Human-Computer Interaction International 2007* (Beijing, P.R. China, July 22-27, 2007). HCI'07. Springer, Heidelberg, LNCS 4551, 1040—1049.
- [3] Epelde, G., et al. 2009. URC Based Accessible TV. In *Proceedings of 7th edition of the European Interactive TV Conference* (Leuven, Belgium, June 3-5, 2009). EuroITV 09.
- [4] i2home. Last access date: 28th January 2009. Intuitive Interaction for Everyone with Home Appliances based on Industry Standards (i2home). <http://www.i2home.org>.
- [5] Vital. Last access date: 28th January 2009. Vital Assistance For The Elderly (Vital). <http://www.ist-vital.org/>.
- [6] Dream Multimedia GmbH. Last access date: 28th January 2009. Dreambox 7020. http://www.dream-multimedia-tv.de/english/products_dm7020.php.
- [7] Microsoft. Last access date: 28th January 2009. Microsoft Vista Windows Media Center. <http://www.microsoft.com/windows/windows-vista/features/media-center.aspx>.
- [8] Carrasco, E., et al. 2007. Enhanced TV for the Promotion of Active Ageing. In *Challenges for Assistive Technology AAATE 2007*. Ed. IOS Press Assistive Technology Research series. Volume 20, pp. 159-163.
- [9] Carrasco, E., et al. 2008. Natural Interaction between Avatars and Persons with Alzheimer's Disease. In *Computers Helping People with Special Needs 2008*. (Linz, Austria. July 9-11, 2008) ICCHP 2008, LNCS 5105, pp. 38-45.
- [10] dot UI. Last access date: 28th January 2009. <http://www.dotui.com/>

An architecture to generate automatic Brazilian sign language legends into Digital Television Systems

<p> Tiago Maritan U. de Araújo Digital Video Applications Lab (LAVID) Federal University of Paraiba Computer Science Department +55 83 3216 7093 maritan@lavid.ufpb.br </p>	<p> Guido Lemos de Souza Filho Digital Video Applications Lab (LAVID) Federal University of Paraiba Computer Science Department +55 83 3216 7093 guido@lavid.ufpb.br </p>	<p> Tatiana Aires Tavares Digital Video Applications Lab (LAVID) Federal University of Paraiba Information Science Department +55 83 3216 7093 tatiana@lavid.ufpb.br </p>
---	---	---

Dênio Mariz Timoteo de Souza
 Digital Video Applications Lab (LAVID)
 Federal University of Paraiba
 Computer Science Department
 +55 83 3216 7093
 denio@lavid.ufpb.br

ABSTRACT

In this paper, we present an architecture solution for improving sign language, specifically Brazilian Sign Language (LIBRAS), in Digital Television Systems. The proposed solution turns possible that deaf people understand TV audio stimulus by using a sign familiar language (LIBRAS). Otherwise, the architecture offers adaptation facilities where regional expressions could be easily incorporated to the LIBRAS legend.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues—Assistive technologies for persons with disabilities D.4.7 [Organization and Design]: Interactive systems

General Terms

Design, Experimentation, Human Factors.

Keywords

Digital TV, Accessibility,

1. INTRODUCTION

Sign languages are visual languages used by deaf people to communicate. Therefore, the different deaf communities need a dictionary that associates signs to the words of the spoken language of their country as well as dictionaries which translate signs from a sign language to another. For example, we have the American Sign Language (ASL)¹, British Sign Language (BSL)², the Italian Sign Language (ISL)³ and the Brazilian Sign Language (LIBRAS)⁴.

Many works are developed for address deaf people needs, these works offer technological solutions for daily activities which enable people with special needs to watch and understand

television, to interact with other people or to write a letter. Describing this scenario we mention the use of emotive captioning in movies and television programs [1] and games for training deaf children [2].

In this paper we approach an architecture that enables the automatic generation of legends in LIBRAS for interactive digital TV programs. The proposed architecture involves a set of software and hardware components integrating broadcasters and digital TV set-top-box.

2. LIBRAS

LIBRAS is the Brazilian Sign Language. It is a visual-gestural language used by the majority of deaf people and recognized by the Brazilian Law No.10436 of 24 April 2002. Additionally we have the standard complement 01/2006 of the Ministry of Communications and the technical standard ABNT NBR 15290:2005. Thus are set rules and guidelines for delivery of audio description, closed-caption and window of LIBRAS in both the analog TV as the digital TV.

LIBRAS is not a simple gesture language, but a way of human expression. LIBRAS language is composed of linguistic levels such as phonology, morphology, syntax and semantics. At the same way that oral-auditory languages, the languages based on signs have lexical items, which receive the name of signals. The only difference is their visual-spatial mode. The communication using LIBRAS is based on a combination of expressions which are the combination of hand movements, points of articulation (local in space or in the body where the signs are made) and facial expressions. Thus, language is a system of transmission of ideas and facts, from communities of deaf people in Brazil. Another important fact is that in LIBRAS there are also regional differences, so you we can find LIBRAS dialects.

The Brazilian government emphasizes that deaf people may use digital television functionalities for translate TV programs into LIBRAS language. In this sense, to issues of accessibility, the use of currently LIBRAS is not restricted to interpersonal

¹ www.lifeprint.com/

² <http://www.kwintessential.co.uk/BSL/history.html>

³ http://elis.eurac.edu/index_en

⁴ www.libras.org.br/

communication, LIBRAS is present in products and services for digital television market.

However, the use of LIBRAS in TV programs is quite limited to manual devices where a window with an interpreter of LIBRAS is shown into the original program. Moreover, there is an operational cost involved to generate this kind of translation. Another problem is that these windows are usually produced at national level not considering the regional features of LIBRAS.

The current digital TV systems do not provide mechanisms for automatic generation of the legend in sign language (as LIBRAS). However, digital TV systems allow the transmission of data flows independent of the TV signal. It is important that TV programs have to be offered with several access options. In some cases, TV programs are the only information sources for people with special needs. And they are not the only benefit. The window with a LIBRAS interpreter, for example, occupies a good part of the screen and can be seen as uncomfortable by most viewers. With digital TV, the window can be optional and resizable (as defined by 01/2006 Standard) so the viewer can choose his best option.

The object of this paper is to describe a system architecture that enables the automatic generation of legends in LIBRAS suggesting an integrated process between the broadcaster and TV receivers. The generated legend will be treated at the receiver as an independent elementary stream (ES). The software of the set-top-box however, can offer features such as enable/disable LIBRAS legend, or resize the window. To address the issue of regionalization and to minimize the computational resources needed to generate the display of the legend in LIBRAS, we propose the use of extended memory (such as USB storage devices) as an alternative to store a dictionary of expressions of LIBRAS language. For example, the dictionary used in the south of Brazil may represent the same expression differently than the dictionary used in northeast. But for the system that is transparent and the viewer has preserved his right to attend the most familiar representation.

3. THE ARCHITECTURE PROPOSAL

In this section we briefly describe the proposal architecture. The main contribution of this architecture is the use of a LIBRAS dictionary at the receptor. The LIBRAS dictionary is responsible for the storage of visual representations related with the sign language. Each sign (or expression) could be represented by an animated image or a video file (as a png or a video file, for instance) and has a code (or index) associated with it representation.

The proposed architecture is shown in Figure 1. Note that at the TV Station scope, we have special components to generate and transmit the LIBRAS grammar. LIBRAS Generator and LIBRAS grammar components generate an elementary stream (ES) composed by a set of LIBRAS codes (grammar). This stream is multiplexed and transmitted with video, audio and data streams to the receptors. At the other hand, the receptor (STB) decodes the LIBRAS elementary stream by associating each code to its representation on the LIBRAS dictionary.

As indicated in Figure 1, on TV Station, a LIBRAS Generator receives an input stream of audio, video or closed caption from a Video Encoder or a Closed Caption Generator. Therefore, using the LIBRAS Grammar, the LIBRAS Generator automatically generates an elementary stream (ES) composed by a set of codes,

where each code is related to a sign (or expression) in the LIBRAS Dictionary. The LIBRAS ES is then multiplexed (or re-multiplexed) on the MPEG-2 Transport Stream (MPEG-2 TS) and transmitted on the Digital TV signal. On set-top-box, there is an extended memory (such as USB storage device) that stores the LIBRAS Dictionary and a LIBRAS Native Application that uses this dictionary to decode and display the signs on screen. As the representation of each sign is stored in the LIBRAS Dictionary, regional aspects are preserved. The native application can offer features such as resize the window, enable/disable LIBRAS legend. This functionality is important for the viewers those could choose the most familiar representation. In this case, we suppose that the set-top-box support to run native applications from USB storage devices as in [3].

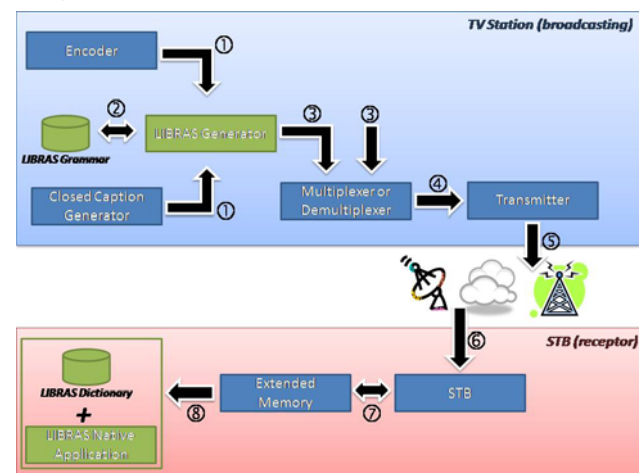


Figure 1 –Architecture Proposal for LIBRAS in Brazilian Digital TV Systems. (1) A/C ES or CC ES (2) LIBRAS grammar (3) LIBRAS ES (4) MPEG-2 TS (5) Digital TV Signal (6) Streams (7) STB Access (8) LIBRAS Dictionary

4. CONCLUSIONS

In this paper, we proposed an architecture solution for improving Brazilian sign language legends in Digital Television Systems. The proposed architecture addresses accessibility features increasing the deaf people interaction into the TV set. Our proposal has an important regionalization requirement that guarantees the LIBRAS dialects preservation. Otherwise, the automatic generation of the legend in LIBRAS is an important innovation function of digital television systems.

5. REFERENCES

- [1] Emotive Captioning. Lee, Daniel, Fels, Deborah and Udo, John Patrick. New York : ACM, 2007. Computers in Entertainment (CIE). Vol. 5, pp. 3-15. Issue 2 (April/June 2007).
- [2] A Gesture Based American Sign Language Game for Deaf Children. Lee, Seungyon, et al. Portland Oregon : ACM, 2005. CHI 2005. p. 4p.
- [3] Souza, Guido Lemos and Leite, Luiz Eduardo. UM COMPONENTE DE COLETA DE INFORMAÇÕES SOBRE SERVIÇOS PARA RECEPTORES DE TV DIGITAL MULTISISTEMA. MU8601656 3 Brasil, 2006. Modelo de Utilidade.

Accessible Digital TV: a research agenda

Giorgio Brajnik
Dip. di Matematica e Informatica
Università di Udine
Udine, Italy
giorgio@dimi.uniud.it

EXTENDED ABSTRACT

The potential of Interactive Digital TV (IDTV) is enormous, given its reachability, the potential ability to provide highly personalized services (when users identify themselves using smartcards, for example to get access to individual e-government services, to health information and services, to social networks), and to act as yet another outlet for using the web.

However these advantages can be hindered by difficulties of using it, including accessibility barriers. Similarly to the web, however, developers of IDTV applications will have no control on the actual devices the user will be using, and will have to address the needs of a very heterogeneous audience. Input devices are likely to be a major hurdle for achieving accessibility. Remote controllers are already very complex devices based on modal interaction and overloading of controls. Furthermore, they support a limited input channel, that poses significant constraints on the kind of applications that can and should be developed.

Another hurdle to overcome is the possible resistance of users to actually use and engage in applications delivered through the IDTV. While over time we should expect that the experience of users in managing digital technologies will increase, during the first years after the switchover, the IDTV will be used also by people that are laggards in terms of technology adoption. Most of them will probably be older adults, with their own accessibility needs. Only if these needs will be met, then IDTV will be able to reach its full potential.

I think that a research agenda on accessible IDTV should include methodological aspects. Accessibility is not only conformance to a set of guidelines¹. While adoption and development of accessibility guidelines is important, it should not be the only aim of the research community. In [3] I discuss some of the shortcomings of an accessibility-as-conformance viewpoint, which we might be able to avoid for IDTV.

I believe the following issues are central to the topic of accessible IDTV.

¹In [2] I listed 9 different definitions. The one I usually prefer in the context of quality assurance is: *accessibility means that disabled people can use the artifact with the same effectiveness as non-disabled people*; this definition, for the web, was first given by [6, 5]

Accessibility Model: Developing a model of accessibility that helps understanding, characterizing and operationalizing the concept. In particular the model should address the following questions: 1. what accessibility is and which properties should be central to it (eg. effectiveness, productivity, usability, satisfaction), 2. which factors influence accessibility and how can they be controlled (eg. experience level with IDTV, attitude towards digital devices, physical conditions under which the IDTV is used, etc.) 3. how accessibility is going to be evaluated and measured (through analytical methods like conformance review, barrier walk-through; through empirical methods like user testing, subjective assessments, etc.).

Accessibility Evaluation Methods: Several AEMs can be used to assess the level of accessibility of an application, but little is known about them. In particular, properties like validity, reliability, usability of Accessibility Evaluation Methods are little studied at the moment. Yet this is crucial to build confidence on methods and obtain a reasonable level of consensus. For the web this did not happen (yet), but more and more researchers are focusing on methods. For the IDTV we will need a set of agile and valid evaluation methods, i.e. 1. such that they can be deployed with limited resources (effort, knowledge, time, infrastructure), 2. that can be scaled up depending on the desired level of quality of their results, 3. that can be used to predict the true problems that will show up when the application will be used.

Database of typical barriers: For practitioners assessing accessibility of IDTV, and for developers as well, it would be useful to be able to draw onto scientific literature presenting evidence of certain kinds of accessibility barriers, and their consequences on users. For the web, specific interpretations of existing (generic and technology-agnostic) guidelines exist, and can be used as a (limited) reference so far. A particular important point here will be the focus on older adults [1, 4].

Sustainable Accessibility: We should be able to provide ways to estimate the return on investment for accessibility

ity practices. For the accessible web, long known and also emerging arguments are based on improved searchability of accessible pages by search engines, on improved ability to search for relevant content in accessible multimedia objects (eg. searching for video scenes using captions), on reduced server bandwidth (because of more compact and better modularized web pages). Unless we can do the same for accessible IDTV, then accessibility projects might not be sustainable, and hence are likely to fail.

References

- [1] A. Arch. Web accessibility for older users: A literature review. W3C, 2008. <http://www.w3.org/TR/wai-age-literature>.
- [2] G. Brajnik. Beyond conformance: the role of accessibility evaluation methods. In S. Hartmann, X. Zhou, and M. Kirchberg, editors, *WISE 2008: 9th Int. Conference on Web Information Systems Engineering – 2nd International Workshop on Web Usability and Accessibility IWWUA08*, LNCS 5176, pages 63–80, Auckland, New Zealand, Sept. 2008. Springer-Verlag. Keynote speech.
- [3] G. Brajnik. Towards a sustainable accessibility. In *Accessible Design in the Digital World*, York, UK, September 22-24 2008. <http://www.addw08.org> York University.
- [4] J. Kim, Y. Pan, and B. McGrath. Personalization in digital television: Adaptation of pre-customized ui design. In *2nd European Conference on Interactive Television: Enhancing the Experience (EuroInteractiveTV 2005)*, Denmark, 2005.
- [5] J. Slatin and S. Rush. *Maximum Accessibility: Making Your Web Site More Usable for Everyone*. Addison-Wesley, 2003.
- [6] U.S. Dept. of Justice. Section 508 of the Rehabilitation Act. www.access-board.gov/sec508/guide/1194.22.htm, 2001.

Effects of Physical Capabilities on Interaction

Pradipta Biswas, Peter Robinson
Computer Laboratory
15 JJ Thomson Avenue
Cambridge CB3 0FD
University of Cambridge, UK
E-mail: {pb400,pr}@cl.cam.ac.uk

Abstract

In this work, we investigated how physical capabilities of users with a wide range of abilities are reflected in their interactions with digital devices. In particular, we investigated the principles of visual perception of visually impaired people and rapid aiming movements of motor-impaired users and also compared those with their able-bodied counterparts. Our studies and results should help interface designers to design inclusive systems and will also enrich cognitive science by explaining the effect of physical capabilities on interaction.

Categories and Subject Descriptors

H.5.2[Information interfaces and presentation]: User interfaces – theory and methods; **K.4.2 [Computers and Society]:** Social Issues – assistive technologies for persons with disabilities.

General Terms

Experimentation, Human Factors, Measurement, Theory

Keywords

Human Computer Interaction, Visual Perception, Hand-strength Evaluation, Rapid Aiming Movement.

INTRODUCTION

One of the basic aims of any interface designer is understanding users. There are many different ways and aspects of understanding users- in this work we investigated how physical capabilities of users are reflected in their interaction with digital devices. Physical capabilities spans through a wide range among users based on age, gender and presence of physical impairments. Lack of knowledge about the problems of disabled and elderly users has often led designers to develop non-inclusive systems. There are guidelines for designing accessible systems (particularly accessible websites), but designers often do not conform to the guidelines while developing new systems. We investigated the principles of visual perception of visually impaired users and motor-action of motor-impaired users and also compared those with their able-bodied counterparts. Our studies will provide the necessary knowledge about the relationship between physical ability and interaction, which will help designers of interactive systems to develop more inclusive systems. We have already used our study to design user models of people with a wide

range of abilities [1]. Our models can be used to determine the optimum font size, contrast and colour of on-screen menu items used to select channels in a digital TV. Similarly, it can also be used to simulate the perception of visually impaired users [2] (like how a person having less visual acuity or colour blindness will view a remote controller) to make designers understand the problems of visual impairment.

ANALYZING PERCEPTION

We designed an experiment to record and analyze eye gaze of visually-impaired and able-bodied users during a visual search task. The task involves searching a shape or icon from a set of distractors. The eye gazes of users were tracked by using a Tobii X120 eye-tracker [4]. Figure 1 shows the average search time for able-bodied users and the same for each visually-impaired user. As expected, the visual search time is greater for visually-impaired users (P1 to P8) than for able-bodied users.

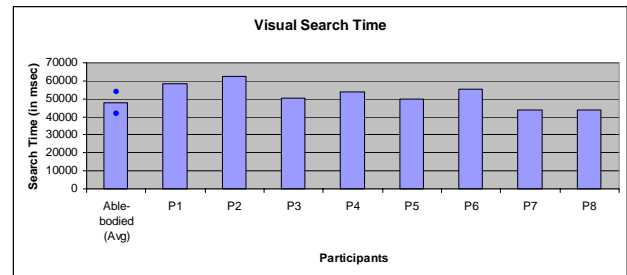


Figure 1. Visual Search Time

If we consider the ‘spotlight’ metaphor of visual perception, a visual search task consists of mainly two steps

- Focusing attention at the probable target
- Moving eye gaze to the next probable target

So we analyzed the details of eye-gaze fixation and eye movement trajectories. We found in the eye tracking data that users often fixed attention more than once on targets or distractors. We investigated the number of fixations with respect to the fixation durations. Figure 2 shows the total number of fixations with respect to the maximum fixation duration. It can be seen that as the fixation duration increases, the number of fixations also decreases. This can be explained by the fact that when the fixation duration is higher, the users can recognize the target and do not need more fixations on it. The number of fixations is smaller when the fixation duration is less than 100

msec, probably these are fixations where the distractors are very different from the targets and users quickly realize that they are not intended target. We also investigated different strategies to explain and predict the actual eye movement trajectory. We did not find any difference in the eye movement patterns of able-bodied and visually impaired users.

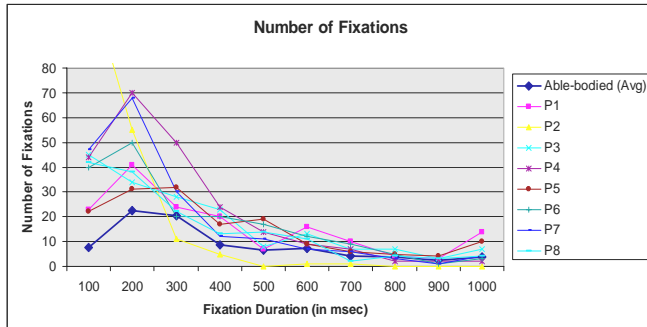


Figure 2. Number of Fixations w.r.t. Fixation Duration

This is due to the fact that the V4 region in brain controls the visual scanning and our visually-impaired participants did not have any brain injury and so the V4 region worked same as the able-bodied users. However visually-impaired users had more number of attention fixations which made the search time longer. The difference between the numbers of fixations for able-bodied and visually impaired users is more prominent for shorter duration (less than 400 msec) fixations. Perhaps this means visually impaired users need many short duration fixations to confirm the recognition of target. From an interface designers' point of view, these results indicates that the clarity and distinctiveness of targets are more important than the arrangement of the targets in a screen. Since the eye-movement patterns are almost same for all users, the arrangement of the targets need not to be different to cater visually-impaired users. However clarity and distinctiveness of targets will reduce the visual search time by reducing recognition time and number of fixations as well.

ANALYZING MOTOR-BEHAVIOUR

In any graphical user interface, a significant portion of interaction consists of pointing tasks. We analyzed and compared pointing performances of motor-impaired and able-bodied users. We made a novel approach of relating hand strength of users with their pointing performance. We found that, for motor-impaired users, the mean and standard deviation of the velocity of pointer movement significantly correlates with the grip strength (Figure 3, $\rho = 0.82$, $p < 0.001$ for mean and $\rho = 0.81$, $p < 0.001$ for standard deviation). We also found that for able-bodied users grip strength and tip pinch strength significantly correlate with the Index of Performance of a 2-dimensional Fitts' Law [3] task ($\rho = 0.57$, $p < 0.05$ for grip strength, $\rho = 0.72$, $p < 0.005$ for tip pinch strength). Our analysis indicates that people having higher hand strength also have greater control in hand movement and can perform pointing faster. Our analysis also showed that flexibility of motion (as measured by the Range of Motions of wrist or fore-

arm) is not as important as strength of hand (as measured by grip strength). We also found similar result for able-bodied users. The positive correlation between index of performance and hand strength shows people with greater hand strength perform pointing faster.

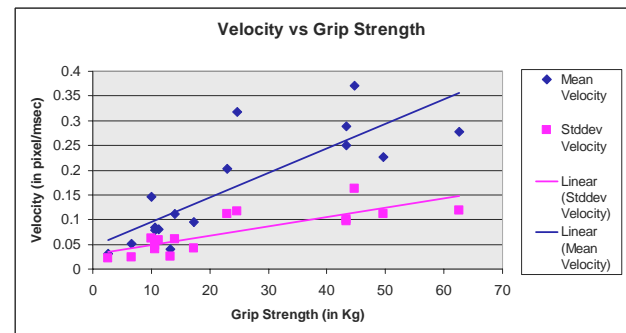


Figure 3. Velocity of Movement vs. Grip Strength

CONCLUSIONS

In this work, we investigated how physical capabilities of users with a wide range of abilities are reflected in their interactions with digital devices. We found that the eye movement patterns are almost same for visually impaired and able-bodied users. However, visually-impaired users fix eye-gaze more number of times than their able-bodied counterparts, which made the visual search time longer. Our study on motor-impaired users indicates that people having higher hand strength also have greater control in hand movement and can perform pointing faster. These studies and results should help interface designers to design inclusive systems and will also enrich cognitive science by explaining the effect of physical capabilities on interaction.

REFERENCES

- [1] Biswas P. and Robinson P., Automatic Evaluation of Assistive Interfaces, ACM International Conference on Intelligent User Interfaces (IUI) 2008, pp. 247-256
- [2] Biswas P. and Robinson P., Modelling user interfaces for special needs, Pradipta Biswas, Peter Robinson, Accessible Design in the Digital World (ADDW) 2008 Available from: http://www.cl.cam.ac.uk/~pb400/Papers/pbiswas_ADDW08.pdf Accessed on: 12/12/08
- [3] Fitts, P.M. The Information Capacity of The Human Motor System In Controlling The Amplitude of Movement. Journal of Experimental Psychology, 47, pp. 381-391, 1954
- [4] Tobii Eye Tracker, Available online: <http://www.imotionsglobal.com/Tobii+X120+Eye-Tracker.344.aspx> Accessed on: 12/12/08

Older Adults and Digital TV

Amritpal S Bhachu
University of Dundee
School of Computing
Queen Mother Building
+44 (0) 1382 388115

abhachu@computing.dundee.ac.uk

ABSTRACT

This paper introduces the PhD research topic of the author and raises questions that are felt to be relevant to the challenges facing the accessibility of DTV. It is the aim that these questions can provoke discussion. It also highlights the areas that the author is interested in as part of the PhD research.

Categories and Subject Descriptors

J.7 [Computers in Other Systems]: Consumer products;

H.5.2 [Information Interfaces and Presentation]: User Interfaces – graphical user interfaces, input devices and strategies, interaction styles, screen design, training, help and documentation, user-centered design.

General Terms

Design, Human Factors, Standardization.

Keywords

Older Adults, Digital Television.

1. INTRODUCTION

From June 2009 I will be beginning my PhD that is entitled, “Adaptive Technologies for Enhancing the Accessibility of Digital TV,” in the School of Computing at The University of Dundee in partnership with BBC Research. The core of this research will be to investigate the range of potential assistive interface technologies that could change the way in which people experience digital television (DTV).

2. BACKGROUND, INTERESTS AND QUESTIONS

The accessibility issues of DTV is an area that is of great interest to me. One group likely to have accessibility issues is older adults. This is a growing demographic of modern day society, largely due to the ‘baby boom’ in the 20th century.

As part of my current project, I have worked closely with older adults to learn more about their thoughts and feelings towards the introduction of home health monitoring, or telecare. This telecare allows their health carers to collect and view data about from the patient’s own home for patients who have chronic diseases. This will allow the carer to find trends in the patient data and allow the patient to remain in their own home for as long as possible, and possibly also reduce the need for hospital visits. It has been established that older adults are the most likely group to have chronic disease. For more information please see the SAPHE website [4].

A workshop was run for this research that involved inviting a group of older adults from the local area to come to the School of

Computing and give their opinions on technology devices that could be used as part of a healthcare system to give them feedback of their own data [1]. This workshop gave me an insight into some of the problems that older adults have with using new technology devices, problems that the designers and developers of these devices must solve. As we prepare for the changeover from analogue to DTV, I see these problems as being relevant to the design and development of DTV, with many questions being raised.

Firstly, it is clear that there is no ‘average’ older user. They have all lived different types of lives and will have different levels of experience when using software based interfaces. Some will be happy to use them and ‘get to grips’ with them with ease, whereas others will have more trouble. Carmichael [2] further illustrates this by showing that although it is accepted that the required software interface abilities diminish with age, the variation between those ability levels also appears to get larger as the target population gets older.

Physical factors that will affect older adults’ usage of DTV include increasingly poor vision, hearing and motor control as well as decreased tactile sensitivity. In the above-mentioned workshop, the problem that many of the group had with the devices was their small screen size and small icon displays. On the devices that had touch screens, the group felt that it could be difficult to ‘click’ on icons as some would find it difficult to be accurate when selecting. This can be easily related to the users having to deal with small icons on their TV set, and their ability to ‘control their remote’.

Necessity of use or perceived value [3] is another of the factors that influences older adults using new technology. During the workshop the group spoke of how younger people would be “seduced by the attractiveness” of the technologies on show [1]. They also felt that older adults were more likely to put “practical usage and cost” above all else. An example of this is the mobile phone. The majority of the group in the workshop had their own mobile phone with a set of basic functions to make and receive calls and text messages. They had adapted to using mobile phones because of the perceived benefits to them i.e. a useful way of keeping in contact with friends and family, as well as in emergency situations. Can it therefore be argued that as television changes from analogue to digital, older adults will find a means appropriate to themselves of controlling their digital television set as they have no other choice? How much of this approach will result in the older adult ‘just getting by’ with using basic functions and not being able to fully embrace the features of DTV? Finally, how much will older adults actually want to use the additional features of DTV?

The ability to learn is another factor that has to be taken into consideration. During the workshop, the older adults felt that having a simple sequence of events when using the devices

would make it easier to remember how to perform tasks. They felt that this process would then become “automatic,” like using a home telephone. A ‘fear factor’ also has to be taken into consideration, as there was a general fear of “breaking” the devices when they used them, and losing the settings of the screen. This again comes into consideration when looking at the menu systems used in digital TVs. Can these menu systems be simplified for use? Can they be customised for individual users, with the settings ‘locked in’? Is it likely that any type of customization would have their own usability issues?

3. FUTURE RESEARCH

The focus of my PhD research will be what the potential assistive interface technologies are and how will they change the way in which people experience DTV. At this stage I am interested in the use of technologies such as Speech Synthesis as a way of rendering text or menus as audio for those who find reading difficult; Speech Recognition to allow users to speak commands to the TV rather than get to grips with complex remote controls; Gesture Recognition used to navigate by pointing a physical device again to move away from the normal remote control and Software Agents to find interesting content to make the amount of material available on DTV easier to handle. However, there are questions before the above areas are addressed of if the core access methods don't work properly for sectors of the population, how will adding technologies help that? Can the basic flow of control and interaction be designed so that it becomes a familiar habit users don't need to think about? There is also the possible impact of making inappropriate choices that have to be considered.

I am also interested in learning more about suitable design approaches when working with older adults. Obvious to this area is the issues raised by using inclusive design and how this will affect the development of DTV and its services. Inclusive design appears to be integral in producing viable solutions for making DTV accessible.

4. CONCLUSION

Through my PhD research, I look forward to becoming part of the process to attempt to solve some of the problems facing older adults using DTV. There are clearly many challenges that lie ahead in finding these solutions, and hopefully many of the questions that have been raised in this paper can be answered during this course of research.

5. REFERENCES

- [1] Bhachu, A. S., Hine, N., and Arnott, J. 2008. Technology Devices for Older Adults to Aid Self Management of Chronic Health Conditions. *Assets 2008 Proceedings*. 59-66
- [2] Carmichael, A. 2002. Older Users of Digital TV – Challenges and Opportunities. *Consumer Policy Review*. 12(6). 198-203
- [3] Rice, M., Carmichael, A. Discovering Older Adults' Perspectives on Undefined TV Applications.
- [4] Smart and Aware Pervasive Healthcare Environment. *SAPHE Homepage*. Online. <http://ubimon.doc.ic.ac.uk/saphe/m338.html>

DEMOS

iFanzzy: A ubiquitous approach towards a personalized EPG

Pieter Bellekens
Eindhoven University of
Technology
Den Dolech 2
Eindhoven, The Netherlands
p.a.e.bellekens@tue.nl

Geert Van Kerckhove
Stoneroos Digital Television
Sumatralaan 45
Hilversum, The Netherlands
geert@stoneroos.nl

Annelies Kaptein
Stoneroos Digital Television
Sumatralaan 45
Hilversum, The Netherlands
annelies@stoneroos.nl

ABSTRACT

iFanzzy is a personalized TV-program guide offering users a helping hand in selecting their ideal television evening in an unobtrusive way. By combining a Web application, a set-top box EPG and a prototype mobile interface, iFanzzy offers a true ubiquitous environment through which users can select and receive personalized TV content. Background information, both data from various heterogeneous sources as well as ontological knowledge repositories, is used to build an integrated RDF/OWL knowledge structure. This semantic graph combining domain knowledge with program metadata is the basis for iFanzzy's main functionality which includes intelligent search and generation of context-sensitive recommendations.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
H.3.3 [Information Search and Retrieval]: Information
Filtering

General Terms

Design, Experimentation

1. INTRODUCTION

The television industry is looking at a fundamental revolution which will change our television experience fundamentally. On the provisioning side, large amounts of television content comprised out of broadcasts, IP channels and user generated content will be available and soon flood currently existing (both electronic as paper) program guides. On the consumer side, the availability of all this content will inevitably lead to the situation where all this content will be made available for various kinds of devices and interfaces (e.g. TV on the mobile phone). These changes bring new possibilities and challenges that affect the whole media chain: from content production, via distribution, to last but

not least the user. With iFanzzy¹, a personalized Electronic Program Guide (EPG) developed in collaboration between Eindhoven University of Technology, Free University of Amsterdam and Stoneroos Interactive Television², we have the ambition to find solutions to the upcoming challenges in fields like Information Retrieval, Human Computer Interaction and Recommender Systems. We believe that via smart personalization strategies we can make sure that a user will always enjoy exactly that content he is most interested in at exactly the time, place and device (like a mobile, a set-top box television combination, a Website, etc.) he chooses. In effect, this leads to the user spending more time with these devices, and in turn gives iFanzzy the ability to constantly closely assess and interpret the user's situation enabling the recommendation of the currently best programs available.

Such an advanced level of personalization among other requires: a rich set of metadata describing all the content, a good understanding of the user's profile including context, feedback and preferences and the semantics of all this data to make a algorithm 'understand' and interpret it. To do so, we rely on a combination of Semantic Web techniques and integration of RDF (Resource Description Framework) and OWL (Web Ontology Language) sources of background information, including both domain ontologies like WordNet³, TV-Anytime⁴, Geonames⁵, FOAF⁶ and OWL Time⁷, as well as external data sources like IMDb⁸ and XMLTV⁹.

Several other recommender systems have been described, each one trying to recommend television programs in a personalized way. In [2] they propose a recommender which looks at predefined stereotypes and preferred user watching times while taking the composition of the current audience e.g. a family into account. Acknowledging the sparsity problem in collaborative recommenders, [3] tries to improve it by using data mining techniques to discover similarities between programs. [1] also uses semantic Web techniques to

¹<http://www.ifanzzy.nl/>

²<http://www.stoneroos.nl/>

³<http://www.w3.org/2001/sw/BestPractices/WNET/wn-conversion.html>

⁴<http://www.tv-anytime.org/>

⁵<http://www.geonames.org/>

⁶<http://www.foaf-project.org/>

⁷<http://www.w3.org/TR/2006/WD-owl-time-20060927/>

⁸<http://www.imdb.com/>

⁹<http://www.xmltv.org/wiki/>

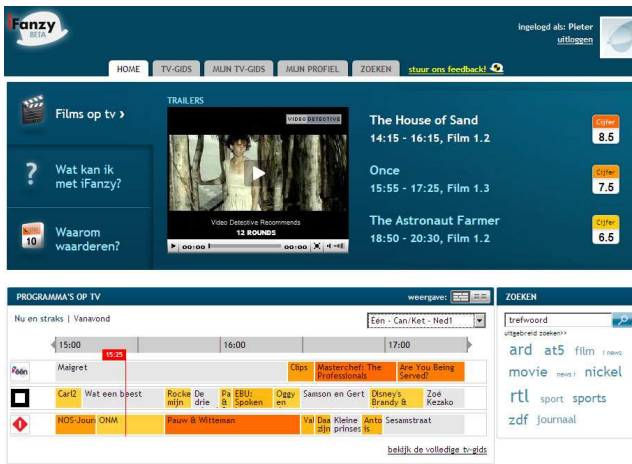


Figure 1: The iFanzy Personalized Program Guide

improve the quality of their recommender system. However we differ in the fact that we combine and integrate information from various heterogeneous live data sources.

2. IFANZY

iFanzy was built to provide a seamless and ubiquitous television experience to the user independent of his choice of platform. Currently iFanzy consists of an Web application, a Personalized EPG or PEPG on a set-top box and a prototype interface on a mobile phone. Each of these platforms is carefully tailored to support the user with the functionality mostly expected from the respective platform. Behind the scenes, all three connect to the same server assuring their mutual synchronization of data. Thereby guaranteeing that every action performed on any of the platforms, has an immediate effect on all. If the user for example rates a program on the Web portal, the set-top box (STB) will take this rating into account when calculating new recommendations.

In figure 1 and 2 we see the interface of the iFanzy Web and set-top box application respectively, both when a user is logged in. In figure 1 we see the home page with in the bottom left a small PEPG listing three channels. The programs shown on these channels get a specific color ranging from very soft toward very strong orange, indicating how well this program fits with the profile of the current user. In this way the interface provides a recommendation for the programs playing right now. For a more elaborate overview, the interface also contains a large listing which can be found under the tab 'TV-GIDS' showing all the channels in the system, colored accordingly in shades of orange. However, the small PEPG differs for everyone as it shows the three most appreciated channels of the current user, either indicated by himself or deduced from his program ratings.

Next to the PEPG, on the right, the front page shows a search bar to look for specific content accompanied by a tag cloud showing the most popular search terms. The iFanzy search takes various semantic relations into account e.g. relations between genres, between words, etc. If a user for example searches for the keyword 'football', all programs with any relation to this specific keyword are found but also



Figure 2: The iFanzy set-top box interface

all programs with a relation to 'soccer' (a hyponym relation in WordNet) or 'basketball' (a sibling ball sport in the genre hierarchy) will be found. The strength of the semantic relation between the concepts determines the ranking of the program in the search results. Furthermore, whenever a program is clicked, a window slides down providing both all the metadata describing this program (including a video if applicable) as well as a set of program functions like "Add to favorites", "Record", "Set reminder", etc. In figure 2 we see that the STB interface is tweaked to match the properties (e.g. resolution) of a television screen. Furthermore, at the bottom of the screen we see that specific iFanzy actions are coupled to the well known colored buttons on the remote control for the user's convenience. A user can log in on the set-top box, by selecting his name and providing a pincode via the remote control. Any new account made via the box also immediately becomes a valid account on the Web.

3. CONCLUSIONS

In this paper we described iFanzy, a multi-platform personalized EPG, which helps the user in finding his way through an ever growing forest of information. By offering a ubiquitous environment which can be accessed from a mobile, a PC or right from the living room, we are confident that users will discover iFanzy's added value in their daily TV watching routine. Currently we have a test running with 50 participants, and another one planned on 500 set-top boxes in Dutch households, to scrutinize our ideas.

4. REFERENCES

- [1] Y. Blanco Fernandez, J. J. Pazos Arias, A. Gil Solla, M. Ramos Cabrer, and M. Lopez Nores. Bringing together content-based methods, collaborative filtering and semantic inference to improve personalized tv. *4th European Conference on Interactive Television (EuroITV 2006)*, May 2006.
- [2] D. Goren-Bar and O. Glinansky. Fit-recommending tv programs to family members. In *Computers & Graphics*, volume 28, pages 149–156. Elsevier, 2004.
- [3] D. O'Sullivan, B. Smyth, D. C. Wilson, K. McDonald, and A. Smeaton. Improving the quality of the personalized electronic program guide. *User Modeling and User-Adapted Interaction*, 14(1):5–36, 2004.

DynamicTV: a new *Inter-tainment* paradigm for television

Rossana Simeoni,
Marina Geymonat,
Roberto Antonini
Telecom Italia Lab
Via Reiss Romoli, 274
10148 TURIN, Italy
+39011228 8041/8765/6459
rossana.simeoni@telecomitalia.it

Roberto Montanari
Department of Science and
Methods for Engineering,
Human-Machine Interaction Group,
University of Modena-Reggio Emilia.
Via Amendola, 2
42100 REGGIO EMILIA, Italy
montanari.roberto@unimore.it

Monica Perrero
Department of Computer Science
University of Torino
C.so Svizzera, 185.
10148 TURIN, Italy
+39011228 8219
perrero@di.unito.it

ABSTRACT

In this demo we propose some of the results of a 3-year project run in the research labs of Telecom Italia, called “DynamicTV”.

It is a new paradigm for iTV, whose drivers, user experience and user interface design are described in [1],[2].

We demonstrate the user experience by a hi-fidelity prototype of the interface, while the system’s end-to-end architecture will be simulated locally. DynamicTV Immersive User Interface allows the user to navigate among media contents simply by using a remote control as it happens in a normal interaction with TV device, albeit a huge number of contents are available, both “on demand” and live on television channels.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces (D.2.2, H.1.2, I.3.6) – *Interaction styles, Graphical user interfaces (GUI), Prototyping, User-centered design.*

General Terms

Design, Experimentation, Human Factors, Verification.

Keywords

Inter-tainment, Interaction Design, Social TV, Long Tail Theory, User Experience, Usability.

1. INTRODUCTION

The DynamicTV project [1][2][3] carried on at Telecom Italia Lab implements a new paradigm for interactive television with the goal of enhancing the user experience, especially in a context where hundreds of thousands of contents are available and the usual zapping or EPG based techniques might not be adequate.

The Inter-tainment paradigm shown in the demo, introduced by the same research group, encompasses and merges different concepts: technologies from enhanced TV, recommendation technologies for supporting the user in the navigation and serendipity. Information visualization techniques have also been

used in order to cope with a huge amount of content to be presented to a user who is usually distant from the screen, therefore not at ease with a lot of textual information, but rather with a visual approach.

2. GOAL

Dynamic TV, is a concrete attempt of implementing the Inter-tainment paradigm approach. This interaction model has to be appealing to both traditional and active users and be shaped in an intuitive navigation metaphor, tailored around users’ fruition habits and familiar interaction patterns. A state of the art analysis of eTV and p2p software applications was led, defining a set of interface design drivers. The aim was to merge the interaction modalities of traditional and widespread entertainment systems with the new approach to content access adopted in new media. As a whole, a new navigation metaphor needed to be defined, to allow different navigation strategies. Possible approaches range between two extremes: on the one hand, the traditional categorical approach recalls the structure of contents arranged in thematic channels; on the other hand, a new serendipitous approach requires contents to reconfigure dynamically whenever the user asks to be advised according to his/her taste, or wants to discover new contents. Dynamic TV’s navigation metaphor is designed to provide users with ways of satisfying the needs that lie behind the two approaches.

3. METHOD

Here are the guidelines we have followed for developing such flexible navigation metaphor:

- The interface design should encourage the “serendipity”[1], i.e. users should be provided with ways to find unpredicted, yet interesting available contents starting from previous search results;
- Users should be provided with suggestions about the available contents, according to a user-targeted recommendation engine;

As counterbalance to the previous guideline, the system should be perceived as non-intrusive, especially as to the provision of recommendations and suggestions.

3.1 User tests

In order to evaluate the appeal and the appropriateness of the user experience with the new TV paradigm, usability and acceptability tests were conducted all along the design process, as part of the design itself.

The initial test were carried on using storyboards, then on low fidelity demonstrators and finally on the prototype shown. The main results are encouraging:

- the prototype was easy to use for every target of users (mean = 3,88 out of 5)
- the appeal of the system was good (mean = 3,86 out of 5)
- users proved to be able to navigate with the usual control device in a simple way (mean usability = 3,86 out of 5) even if not always efficiently

3.2 Demonstrating the paradigm

The demonstration will show side by side two versions of Dynamic TV: this is important in order to show that we are not demonstrating only a User Interface, but a whole “paradigm”.

The principle of the paradigm are very well represented by a basic (called “Zen”) version, represented in Figure 1, and are even more apparent when comparing it with a graphically skinned version, where the graphics are very different, but the User Experience and Interaction Rules are exactly the same.

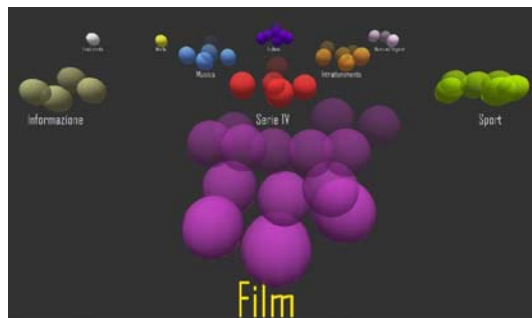


Figure 1 – Dynamic TV “Zen” appearance

The latter version, called “Space” and shown in Figure 2, is demonstrated together with the “Zen”, both to underline the characteristic features of the paradigm across different graphical layouts, and to show how the software architecture allows to perfectly decouple User Interaction from Graphical aspects.



Figure 2 – Dynamic TV “Space” appearance

3.3 What’s next?

The project is currently focused on including in the prototype some cross-media features described in [4]. The idea is to leverage on the actions that a user from his PC does on the iDynamic TV website, such as voting, tagging, using social networks to share comments, etc, in order to enrich the user experience in the fruition of television contents, adding a participation and sociality dimension.

4. TECHNICAL DESCRIPTION

The client shown runs on a normal PC with media centre features. It is a web based application composed by two layers:

- a presentation layer, which provides the User Interface both immersive 3D navigation and multimedia fruition, the latter through a pluggable media player
- an application layer, where all client-side business logics take place

plus a “skin” storage, which allows the graphics and scene structure to be changed without modifying the paradigm at all.

The client is part of an end-to-end architecture, including an application server, providing information to build the 3D scene, a metadata DB, a content delivery network, which supplies VoD and live TV to the customer, and a recommendation engine.

5. ACKNOWLEDGMENTS

Maurizio Belluati, Paolo Garbin (TILab) for prototyping. Elena Guercio (TILab), Francesco Tesauri (University of Modena and Reggio Emilia) and Amon Rapp (University of Turin and TILab Lagrange Project) for user tests and design. Agnese Vellar (Universita’ of Turin): for social studies and design.

6. REFERENCES

- [1] Simeoni, R., M. Geymonat, E. Guercio, M. Perrero, A. Rapp, F. Tesauri and R. Montanari (2008). Where Have You Ended Up Today? Dynamic TV and the Inter-tainment Paradigm In the Proceedings of the 6th European Conference, EUROITV 2008, Salzburg, Austria, Springer Berlin / Heidelberg, pp. 238–247.
- [2] Simeoni, R., Etzler, L., Guercio, E., Perrero, M., Rapp, A., Montanari, R., Tesauri, F.: Innovative TV: from an old standard to a new concept of Interactive TV – an Italian job. In Julie A. Jacko (Ed.): Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, 12th International Conference, HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part III. Lecture Notes in Computer Science pp. 971-980, Springer (2007).
- [3] Belli, A., Geymonat, M., Perrero, M., Simeoni, R., Badella, M. DynamicTV: the Long Tail applied to broadband-broadcast integration, Euro ITV 2006.
- [4] Carmagnola, F., Cena, F., Console, L., Vernerio, F., Simeoni, R., Cortassa, O., Perrero, M. “iDYNamicTV: Web 2.0 and adaptation for a new television experience”, UXTV2008

Automatic Program Extraction From TV Streams

Gaël Manson

Xavier Naturel

Sid-Ahmed Berrani

Orange Labs – France Telecom
Division R&D – Technologies
4, rue du Clos Courtel
35510 Cesson Sévigné. France

{gael.manson, xavier.naturel, sidahmed.berrani}@orange-ftgroup.com

ABSTRACT

This demo presents a method of detecting television programs in TV streams. The objective is to automatically detect the precise start and end of broadcasted television programs. The method first detects inter-programs (commercials, trailers...) as repeated sequences in the broadcast stream, and then deduces the programs boundaries. The extracted programs can then be stored in a database to be used in novel services such as TV On Demand. The demo shows examples of repeated inter-programs detection and program extraction, performed on several days of French television from two different channels.

Categories and Subject Descriptors

H.3.1 [Information Storage and retrieval]: Content Analysis and Indexing; I.4.9 [Image Processing and Computer Vision]: Applications

Keywords

Video indexing, video processing, commercials detection.

1. INTRODUCTION

TV On Demand services aim at making TV content available to viewers without any constraint of location and/or time. Viewers have the possibility to access past TV programs from a large range of channels. To build these services, a way of extracting individual programs from television streams is needed. One of the solutions used nowadays, is to perform this extraction manually. This is not efficient and may prove difficult and costly to achieve, when a large number of programs and channels is considered.

Metadata, like Event Information Table (EIT) or Electronic Program Guide (EPG), provide information on the structure of TV streams. They are unfortunately imprecise, incomplete and not always available [1]. Moreover, TV channels cannot usually provide accurate metadata because of technical limitations, especially the complexity of the audio-visual chain. For historical and organizational reasons, services making use of TV content are also generally developed by third parties [2].

This demonstration is based on the novel techniques proposed in [2, 3, 4]. It shows how these techniques can be put together in order to perform accurate and fully automatic TV program extraction.

2. OVERVIEW OF THE SYSTEM

TV programs are heterogeneous and do not share any common features. It is therefore very difficult to detect them directly. However, Inter-Programs (IP), e. g. commercials, sponsorship, trailers, self-advertisements... are generally broadcasted several times a day in the stream. IP can thus be automatically detected as near-identical repeated sequences, and programs can be deduced as the rest of the stream. This is a very generic solution that can be applied regardless of the channel or the country, and does not make any assumptions about broadcasting rules (e.g. black frames...). This is the basic principle of our approach. Repeated sequences are first detected. They are then classified and IP are identified. Programs are then finally deduced.

This demonstration presents results on TV streams from two French channels (one public, one private). The demonstration is composed of two modules: the first one shows how inter-programs are detected; the second one shows how programs are extracted.

2.1 Module 1: inter-program detection

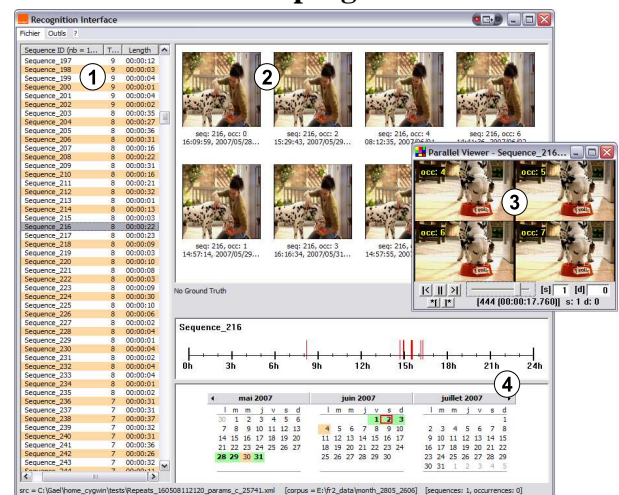


Figure 1: Near-identical repeated audiovisual sequences. (1) Detected repeated sequences. (2) Occurrences of the selected repeated sequence. (3) Some of the occurrences played in parallel. (4) Date and time of the occurrences.

IP are detected by their repetition property. The method identifies repeated sequences from a clustering-based approach that groups similar keyframes [2]. Similarity of keyframes is computed on a visual descriptor built from quantized DCT coefficients [2]. Repeated sequences are then used to segment the stream: each occurrence of a repeated sequence is considered as a segment and each gap between two consecutive segments is also a segment. Each of the resulting segments could, hence, be part of a program or of an inter-program. Therefore, these segments have been classified based on their intrinsic features (e.g. duration, number of repetitions) and on the relational and contextual information of the repeated sequences as explained in [3].

Figure 1 shows near-identical repeated audio/video sequence detection results. The set of detected repeated sequences can be easily browsed, and it can be checked that all detected sequences are actually repeated sequences: (1) all detected repeated sequences can be viewed with their mean duration and their number of repetitions, (2 & 3) several occurrences of one repeated sequence can be played in parallel in order to compare their content and their context and (4) all occurrences are placed on a calendar to show their frequency.

Through the same interface, it is also possible to browse the whole set of resulting segments, or to focus on segments classified as inter-programs.

2.2 Module 2: TV program extraction

The previously classified segments are used to segment the stream. Consecutive segments that have been classified as inter-programs (resp. programs) are merged. The resulting program segments are then labeled using a straightforward matching procedure using the available metadata. The programs are finally extracted.

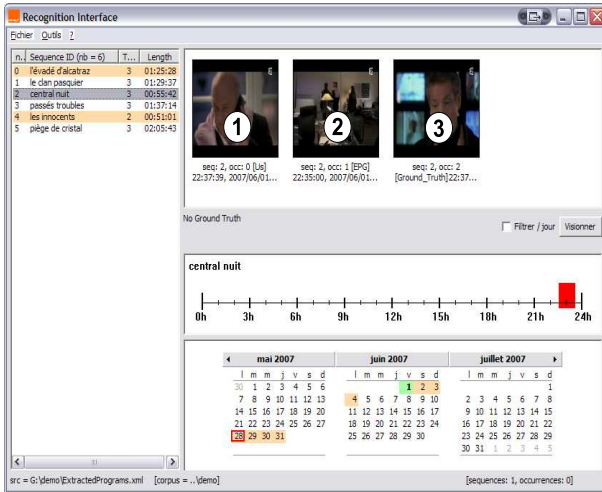


Figure 2: Extracted TV programs. (1) Our program extraction. (2) EPG-based program extraction. (3) Ground-truth-based program extraction.

Figure 2 shows extracted TV programs. Results (1) are compared to the EPG (2) and to the ground-truth (3). This interface allows us to see that our solution is very accurate. All the programs can be played in parallel. In [3], an evaluation study of our programs extraction solution is presented on 18 hours of TV streams. The obtained results are summarized in Table 1. This table presents the mean of the absolute values of the differences between the obtained

start (resp. end) times and the ground-truth start (resp. end) times of programs. This table also provides the accuracy of an EPG-based program extraction.

	Program Start time	Program End time
Our Solution	5.6 s	11.7 s
EPG	2 m 14.0 s	4 m 6.5 s

Table 1: Accuracy of program extraction.

Overall, these results show that the proposed method of program extraction is very accurate and outperforms the metadata.

3. CONCLUSION

This demo shows how the modules of our automatic TV program extraction system work. It performs automatic extraction and labeling of TV programs from a television stream. It also allows us to explore results obtained on real TV streams of two representative French TV channels. Future extension will consist in building an online system, capable of extracting programs on the fly.

4. REFERENCES

- [1] BERRANI, S.-A., LECHAT, P., AND MANSON, G. TV broadcast macro-segmentation: Metadata-based vs. content-based approaches. In *Proceedings of the ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands* (July 2007).
- [2] BERRANI, S.-A., MANSON, G., AND LECHAT, P. A non-supervised approach for repeated sequence detection in TV broadcast streams. *Signal Processing: Image Communication, special issue on "Semantic Analysis for Interactive Multimedia Services"* 23, 7 (August 2008), 525–537.
- [3] MANSON, G., AND BERRANI, S.-A. An inductive logic programming-based approach for TV stream segment classification. In *Proceedings of the IEEE International Symposium on Multimedia, Berkeley, CA, USA* (December 2008).
- [4] MANSON, G., AND BERRANI, S.-A. Tv broadcast macro-segmentation using the repetition property of inter-programs. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications, Innsbruck, Austria* (February 2009).

Fernweh - Prototype of an IPTV-Community for Travel and Language

Bjoern Muenchau
meta.morph GbR
Fidicinstraße 3
10965 Berlin, Germany
+49/(0)30/202 35 925
muenchau@meta-morph.tv

Robert Strzebkowski
Beuth Hochschule für Technik
Luxemburger Straße 10
13353 Berlin, Germany
+49/(0)30/4504-2013
robertst@tfh-berlin.de

Jonas Neumann
meta.morph GbR
Fidicinstraße 3
10965 Berlin, Germany
+49/(0)30/202 35 925
neumann@meta-morph.tv

ABSTRACT

This paper outlines a prototype developed to demonstrate a casual learning and social meeting platform for travel, countries and languages in an IPTV environment.

Categories and Subject Descriptors

K.3.1 [Computer and Education]: Computer Uses in Education - Distance learning

H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces - Computer-supported cooperative work

Keywords

t-learning, casual learning, social platform, travel, languages, countries, community, individualization, personalization, IPTV

1. BACKGROUND

The audience expects more than passive reception of TV content from digital television. Among other investigations, a survey about BSKyB's well-established interactive TV services shows that about 60% of all customers already used the applications, 40% regularly and intensively [1].

Most viewers use the offerings a) to request additional information about the currently watched program, b) to customize the program to individual interests and c) to receive information about commercial products that currently appear. So mainly for services which satisfy the need for further information and personalization.

The same survey points out that viewers who already gathered experience with digital interactive media, e.g. user of VoD services or VCRs with time-shift features, are much quicker thrilled by other interactive services. A current study about German TV reception also shows an increase of interest in information and infotainment content [2].

These key facts essentially point to an increased users' interest in challenging interactive scenarios, especially in the field of information and infotainment.

In addition a better part of the target group also uses the Internet frequently, thus has experience in handling interactivity and modern cross-linked offerings. Web 2.0 services have become more and more popular over the last years and the vast user majority use collaborative websites regularly. A huge group of frequently used Web 2.0 applications are social platforms, e.g. communities and forums [3].

Another study shows that "travelling" (right behind "watching TV" and "meeting friends") is a very important topic regarding German leisure behavior [4]. That is one reason why documentaries about expeditions and travel programs rank

among the most favorite TV contents [5]. After "health", "learning languages" was the second most important educational topic in Germany in 2007 [6].

According to a worldwide survey "learning a language" and "developing basic reading and computational skills" are interactive services that could be delivered to any household connected to digital television. 95 to 99 percent of all European homes own a TV set, therefore "t-learning" (learning with digital interactive television) could become an important method and open new ways for homelike learning in the near future. Another survey result is that there is a middle- to long-term potential in the use of digital television for distance learning, especially with personalized services [7].

Adequate to the possibilities of interactive television, with a remote control as the only input device, "t-learning" is mainly not about complex issues or audit proof facts but about aspects of motivation, stimulation, repetition, short exercises and support at finding solutions and information. Learning should take place in a pleasant, exciting and (TV) conformal atmosphere.

Media didactics research proves that learning becomes interesting, is enjoyable and effective if the learning process a) has a real value for the learner, b) is closely connected to a real situation, c) actively involves the learner and d) allows the exchange with other learners or the teacher [8].



Figure 1: Main menu and entry portal

2. FERNWEH

Fernweh is a concept and prototype for an interactive TV application to meet the described interests in travel, languages and information/infotainment content. The audience can gather customized data about geographical regions, go on virtual journeys and participate in exciting language exercises, all at the same time and within the same user interface.

At first the user edits a personal profile (e.g. name, age, sex, languages, favorite destinations, countries of interest) and receives a personalized entry portal to the relevant topics – language, travel and community.

2.1 Language

With a short optional test the personal language skills can be determined. The viewer browses through various multimedia exercises and receives a matching recommendation. Either a serious learning path can be taken or playful challenges lead to the aspired goal. Even if learning is only an entertaining matter, solving the exercises continuously leads to an improvement of the language level, but at the end it is always the learner's ambition that decides about the level of success.



Figure 2: Language menu with several modules

The learning paths use short video clips, pictures, audio files as well as (simple) interactive games to tap the full potential of the medium. Strong situational elements (e.g. daily experiences, funny short stories related to a specific region) should motivate the audience by personal interest and bind in the long term.

2.2 Travel (and Countries)

The application also delivers content that has a direct connection to travel and countries. The viewer can search for certain places or regions and receives initial suggestions that match previously gathered interests. Thereby every user gets an individual "virtual" program.

In addition, the area "Travel" gives access to user generated travel reports from a connected community and travel enthusiasts can share their personal experiences, find interesting travel groups or plan future journeys together.

2.3 Community

Beside its formal connection, "Travel", "Language" and "Countries" are linked with several community features. The users' profile is the administrative centre for all personal data and all deposited information influences the "Member" area, a communication hub to other interesting members. Viewers can exchange personal information or find learning/travel buddies here.

In addition there are several services that are directly related to travel or language. Among other offerings the user has direct access to news, weather and country specific information or can look for interesting travel groups, always depending on the users' profile.



Figure 3: Community menu to join a travel group

3. CONCLUSION

To gain the aspired interconnectivity of services and achieve a high level of individualization and personalization, it is necessary that providers of audiovisual and print-based contents (e.g. media houses), educational institutions and travel operators work hand in hand, to create a uniquely new format for interactive learning and virtual travelling.

4. REFERENCES

- [1] De Freitas, S. & Bagur, B., 2007: A case study analysis on digital convergent design: SkyNet platform. In: Lugmayr, A. & Golembiowski, P. Interactive TV: A Shared Experience, TICSP Adjunct Proceedings of EuroITV 2007
- [2] Gerhards, M. & Klingler, W., 2007: Mediennutzung in der Zukunft, Media Perspektiven 6/2007, pp. 295 – 309
- [3] Gscheidle, C. & Fisch, M., 2007: Das „Mitmach-Netz im Breitbandzeitalter, Media Perspektiven 8/2007, pp. 393 – 405
- [4] IFAK - Institut Markt- und Sozialforschung, 19. Oktober 2006: Demographische Veränderungen -welche Anforderungen stellen sie an das Marketing der Zukunft? Document of 20th Medientage München
- [5] Zubayr, C. & Gerhard, H., 2006: Tendenzen im Zuschauerverhalten, Media Perspektiven 3/2006, pp. 125 - 137
- [6] Statistisches Bundesamt, 2007: Weiterbildung. Wiesbaden, <https://www.ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1021199>
- [7] Bates, P.J., 2003: t-learning Study, A study into TV-based interactive learning to the home, pjb Associates
- [8] Issing, L. J. & Klimsa, P., 2004: Informieren und Lernen mit Multimedia, Weinheim, BELTZ/PVU Verlag

MyMedia: Dynamic Personalization of Multimedia

Alexander Voß

Microsoft EU Innovation Center

Ritterstraße 23

52072 Aachen

+49.241.99784.34

avoss@microsoft.com

Chris Newell

BBC Research & Development

Kingswood Warren

Tadworth, Surrey KT20 6NP, UK

+44 1737 839659

chris.newell@rd.bbc.co.uk

Paul Marrow

BT Innovate

Orion 3 Room 7 PP1

Adastral Park, Ipswich IP5 3RE, UK

+44 1473 641456

paul.marrow@bt.com

ABSTRACT

This paper describes a demonstration by the MyMedia EC-funded collaborative research project at the 2009 European Interactive TV conference. The MyMedia project addresses the key social problem of information overload that has been called the "Crisis of Choice". The project has developed a recommender system which provides personal recommendations about TV and radio programs tailored to specific individual tastes and interests by analyzing implicit and/or explicit user feedback.

This demonstration will show the current state of development of the recommender system technology and the associated user interface design work as the project prepares for four independent field trials.

Categories and Subject Descriptors

H.3.3 [Information Storage And Retrieval]: Information Search and Retrieval – *Information filtering, Relevance feedback, Selection process.*

General Terms

Algorithms, Design, Human Factors.

Keywords

Personalization, Recommender, Social Networks.

1. INTRODUCTION

The MyMedia project has created an open source software framework which allows the integration of multiple, Audio/Visual content catalogues and recommender algorithms into a single system. Application developers that want to incorporate the Dynamic Personalization Framework in their applications can easily use the architecture and tools provided by the project to create a recommender engine.

The framework architecture provides a platform for the evaluation of recommender algorithms [1] and for exploring metadata enrichment technologies [2, 3]. The project is also exploring the

creation of media-centric social networks to exploit the potential of shared recommendations and user generated metadata such as tags [4].

2. FIELD TRIALS

The MyMedia project includes four comprehensive field trials in three different countries. Each trial will provide a user interface through which it will gather implicit and/or explicit user feedback.

The BBC field trial in the UK will explore MyMedia in the context of its developing online catch-up services. It will study the specific difficulties raised by continuous broadcast services where there is a constant stream of new content for which there is initially no user feedback for recommenders to exploit.

The BT field trial will investigate the provision of MyMedia recommendation services through its BT Vision IPTV Video-on-Demand purchasing service. Viewers will be offered recommendations to view and the trial will investigate their responses in terms of experience of the recommendation service, and effect on purchase of Video-on-Demand content.

The Microsoft field trial will investigate the impact of personalized recommendation services on the end user perceived usefulness of Microsoft's MSN Video content offerings. The field trial will incorporate social networking technology and analyze if recommendation services are influential to the personal relationships and media consumption habits of people in a given application context. Furthermore it will test novel ways to present content offerings and incorporate user feedback and analyze how they affect the user experience.

The Microg nesis field trial in Spain will involve Yoguie, an e-commerce application for audiovisual and musical content.

3. DEMONSTRATION

3.1 BBC demonstration

The BBC demonstration will show the use of the MyMedia Framework to provide personalized recommendations in the context of an online 7-day catch-up service for viewers of the BBC's broadcast TV and radio services.

The user interface will allow viewers to contribute explicit feedback about content and recommendations whilst also collecting implicit feedback. The benefits of personalized recommendations will be shown in comparison to non-personalized recommendations based solely on similarity and context.

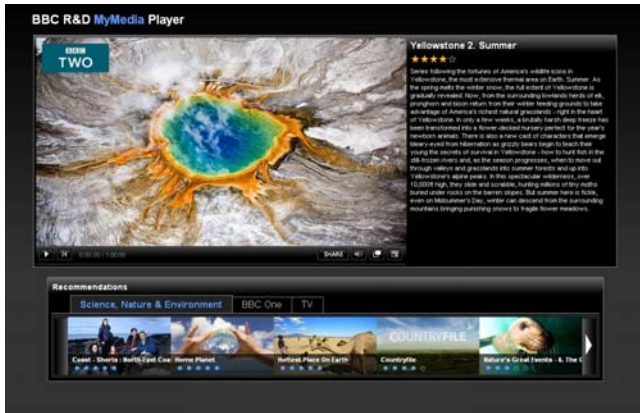


Figure 1. BBC recommender demonstration.

3.2 BT demonstration

BT's field trial will be held in the UK in households and lab conditions with IPTV services provided through BT broadband. The BT demonstration will show how viewers can interact with the MyMedia recommendation toolkit in a novel IPTV context, through a simulation of this user experience generated from a laptop.

Viewers of the demonstration will initially see the front screen of the BT Vision service, and will be able to browse from this screen to other screens of the service, and see how recommendations provided by the MyMedia recommendation toolkit can be seamlessly integrated with the service.

Viewers can find out more information about recommended items, choose whether to purchase or not these recommendations, or follow up other directions. As viewer interaction with the system develops, the user profile should become more sophisticated and recommendations should correspond better to user expectations.

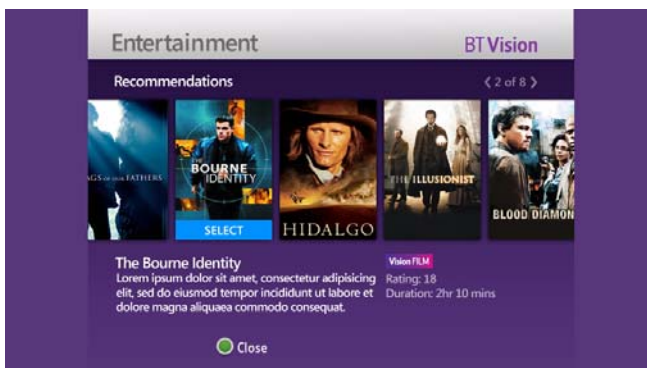


Figure 2. BT recommendations via BT Vision.

3.3 Microsoft demonstration

Microsoft will show a feature rich live demonstration of a prototype of MSN ClipClub which is a video sharing and social

networking application that includes a video recommender based on the MyMedia recommender framework.

The demonstration will start with an overview of the novel user interface that enables people to interact with the system and provide feedback of a user's preference in a seamless fashion.

It will show in the live demonstration that even with few user feedbacks the MyMedia recommender framework can generate highly personalized media recommendations. Users are enabled to identify other people with similar likings based on their preferences and provided with an interface to share content offerings, exchange instant messages or invite people to the user's social network. Users can monitor and control their privacy settings to decide what preferences are shared with a broader public community.

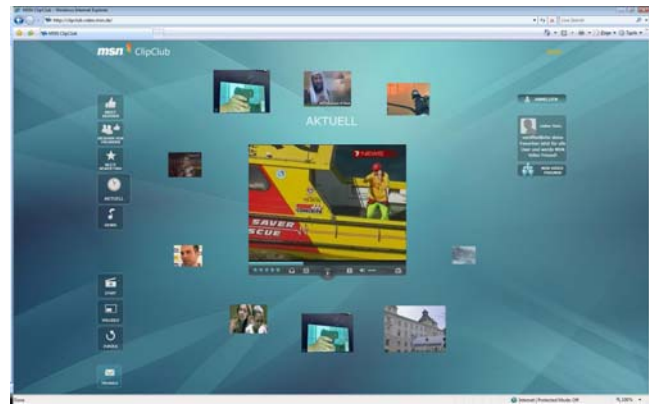


Figure 3. MSN ClipClub demonstration.

4. ACKNOWLEDGMENTS

MyMedia is a collaborative research project funded by the European Commission. The partners are: European Microsoft Innovation Center, BT Research, BBC Research & Development, University of Hildesheim, Telematica Instituut, Technical University of Eindhoven and Microgenesis.

5. REFERENCES

- [1] Rendle, S. and Schmidt-Thieme, L. 2008. Online-Updating Regularized Kernel Matrix Factorization Models for Large-Scale Recommender Systems. ACM RecSys. (Lausanne, Switzerland, Oct. 23-25, 2008).
- [2] Wartena, C. and Brussee, R. 2008. Topic Detection by Clustering Keywords. TIR 2008 Workshop of DEXA2008 (Politecnico di Torino, Italy, Sept. 1-5, 2008).
- [3] Wartena, C. and Brussee, R. 2008. Instance-Based Mapping between Thesauri and Folksonomies. International Semantic Web Conference (Karlsruhe, Germany, Oct. 26-30, 2008).
- [4] Bollen, D. and Halpin, H. 2009. The Role of Feedback in Folksonomies: How Social is Social Tagging? 18th Int. World Wide Web Conference (Madrid, Spain, April 20-24, 2009).

Discot: A Platform to Approach Information Society to Everyman Users

Alex Ugarte
VICOMTech
Paseo Mikeletegui 57
20009 San Sebastián
Guipuzcoa-Spain
+34 943 309230

augarte@vicomtech.org

Julián Flórez
VICOMTech
Paseo Mikeletegui 57
20009 San Sebastián
Gipuzkoa-Spain
+34 943 309230

jfllorez@vicomtech.org

Igor G. Olaizola
VICOMTech
Paseo Mikeletegui 57
20009 San Sebastián
Gipuzkoa-Spain
+34 943 309230

iolaizola@vicomtech.org

David Oyarzun
VICOMTech
Paseo Mikeletegui 57
20009 San Sebastián
Gipuzkoa-Spain
+34 943 309230

doyarzun@vicomtech.org

ABSTRACT

Information society offers the possibility to access content and services, to make business, to communicate people. However not everybody have the skills required to use a PC in order to access these new services. The group of users that are let out of this information society is called the digital gap. On the other hand, we have television sets, a platform with a huge penetration that everybody is capable of using, but with not so many applications.

Actions taken to reduce the digital gap, usually imply trying to make PCs more accessible to everybody. In this paper, in the contrary, we propose an approach where applications are moved into the television environment, so that users do not require new skills to interact with this new content. In this paper we present the architecture of an advanced media center that makes this approach possible.

Categories and Subject Descriptors

H.5.1 [Multimedia Information System]

H.5.2 [User Interfaces]

General Terms

Design, Experimentation

Keywords

Advanced media center, Digital gap, Interactive TV, smart homes, Internet TV

1. INTRODUCTION

Advances in personal computers and Internet connections have made possible what is now known as Information Society. However, not everybody is included in this digital society. In order to have access to these new contents, it is necessary to have access to Internet and some skills with computers. Although the

possibilities offered are huge, the number of persons that have access to them is limited. This is known as the digital gap.

The traditional way to reduce the digital gap is adding "accessibility options" or adding new items to the different menus that should do things easier. In [1] a new approach is presented in which software complexity is reduced to offer just the functionality required by unskilled users. In this work we go a step further; in addition to maintaining the interface simple and giving very specific functionality, we think that changing the platform in which the services are offered is crucial. People are afraid of the PC, but they feel comfortable with the television.

Moreover there is still an interest in the success of television as a platform to execute applications. Governments see it as a good way to offer services to everybody. The fact that almost 100% of the population has access to a television set and 80% of the population has a telephone line to be used as a return channel makes this platform a very attractive way to offer e-government and e-health services [5].

A lot of research has been done on how to design interfaces for television. It is well known by anyone who designs applications for television that it is a complete different world from PC [2]. Not only the attitude of the user is different, as in the television he is more relaxed and passive, known as lean-back attitude, but also the characteristics of the monitors are also different. Thus, the graphical design must be done taking these facts into account [3] [4].

With the lack of massive deployment of standard broadcasted applications (deployment of MHP was a failure in Europe), the responsibility to fulfill all these expectations of the television being a platform for interactive applications relies on media centers. Devices as MythTV, Tivo and Windows Media Center are popularizing the concept of having a computer connected to the television. These media centers, which up to date are centered on media content, in combination with the concept of smart homes, intelligent devices and the protocols related to them (such as UPnP: Universal Plug and Play [7] and OSGi: Open Services Gateway initiative [8]), provide the perfect environment on top of which a set of useful and easy to use applications for everyman user can be built.

In this demo we present Discot, a media-center-like software that provides the adequate platform to offer new applications and services on the television. In contrast with other media centers that focus in the media content, and delegate other tasks to the operating system, Discot aims at offering a full solution for set-top boxes, trying to depend as less as possible on other



Figure 1. Two examples of scene composition. On the left the avatar is scaled to one quarter of the screen area. On the right the television widget is scaled to keep the focus on the browser.

applications available on the underlying operating system. The platform tries to use as many standard protocols and specifications as possible to make it compatible with an horizontal market.

2.SYSTEM DESCRIPTION

Two applications have been integrated: a web browser and an avatar. The web browser offers the possibility to present already existing information on the television. Even though the web browser is capable of rendering HTML and executing some JavaScript code, existing web content is not adapted to be presented on television, and therefore the browser should not be considered as a general purpose Internet browser, but as tool to access specific adapted portals. The avatar will offer a multi-modal human computer interface [6] that complements the traditional menu-based system.

More functionality can be added by providing new plug ins to the platform. The API available to develop new plug ins, automatically makes them controllable with UPnP protocol, and offers an easy way to compose the scene with the TV screen.

3.Demo Scenario

The scenario where this demo takes place is in a relatively intelligent environment, i.e. a house with a home network, with some smart devices interconnected. In this intelligent environment the TV would be used as the interface to interact with the system.

In this scenario we will present two different use cases. A screen shot of each of them can be seen in Figure 1.

♦**Doing exercises:** In the first use case, the user actively initiates an interaction with a service provided in the platform. In this case the user is an elderly person to whom the doctor has told to do some exercises to keep fit. But this time, instructions to do the exercises are not given in a paper with some images, but he can follow a virtual monitor (an avatar) on the TV that does the exercises at the same time as he does. With a text to speech synthesizer, the avatar gives adequate instructions, and it is also possible to select different point of views.

♦**Supervising home status:** The user has configured the smart home to keep some comfort values. For example it has been set to



keep the temperature over 20°C. However, and even the central heating is on, the temperature keeps going down and now it is on 19°C. The central systems analyses the information from all sensors in the house and detects that a window is opened. It then presents a warning on the TV indicating this situation. The user then can close the window. Other kind of notifications are possible in this scenario. Note however, that the demo presents the possibility to use the TV as an interface for the whole system, but that the sensors and the intelligence of the system is emulated in an OSGi framework.

4.REFERENCES

- [1] Gregor, P. and Dickinson, A. Cognitive difficulties and access to information systems – an interaction design perspective. *Accessibility and Computing. A Regular Publication of the ACM Special Interest Group on Accessible Computing*. Number 83 (Sept 2005), 59-63.
- [2] Lekakos, G, Chorianopoulos, K. and Spinelis, D. Information systems in the living room: a case study of personalized interactive TV design . *Global Co-Operation in the New Millennium. The 9th European Conference on Information System*. Bled, Slovenia, June 2001
- [3] Darnell, M.J. Effects of Font Style, Size and Color on the Readability of Text on TV. *Proceedings of EuroITV, 2005*. 125-131.
- [4] Flórez, J., García, I., Aizpurua, I., Paloc, C., Ugarte, A., Agirre, M., Colet, J. and Zubiaur . SEITV- Interactive multimedia leisure/educational services for Digital TV in MHP. *Entertainment Computing - ICEC 2004, Third International Conference, Eindhoven, The Netherlands, September 1-3, 2004, Proceedings*.
- [5] Rascioni, G. Interactive Applications for the Public Administration in Digital TV Environment. *Adjunct Proceedings of EuroITV 2008*.
- [6] González, M.F., Buiza, C., Etxeberria, I., Yanguas, J.J., Ortiz, A., Carretero, M.P., Oyarzun, D. and Garay-Vitoria, N. The effect of an avatar in natural interaction. *AAATE07: Challenges for Assistive Technology*, pp.893-898. Donostia - San Sebastián, Spain, 2007
- [7] UPnP Forum: <http://www.upnp.org/>
- [8] OSGi Alliance: <http://www.osgi.org/Main/HomePage>

iTV IN INDUSTRY

Designing Alphanumeric Input for iTV in a Search Context

AIM OF THE STUDY

One of the prime implications of networking and socializing an interactive TV system is the growing need for an efficient user input interface. We aimed to use specifically relevant in-house academic and industrial research experience to elicit market insights on an alphanumeric input method we designed for a digital television service offered by a cable provider in Quebec, Canada. After substantial research and design, our remote-controlled search interface was implemented and is currently being used by real users.

The aim of this study was to revisit and improve conclusions drawn from our initial research with real users of a real product. The final paper/presentation demonstrates the initial design challenges, traces the project stages, outlines the basic usability principles which emerged, and presents how those principles evolved after retrospective user research.

METHODOLOGY

Quantitative User Research: recording users' performance and efficiency while engaging in controlled interaction with the interfaces (success rates, user paths, questionnaires, etc.).

Qualitative User Research: interviews and focus-group studies of real service users. Data on subjective user experience, intentions, motivations, attitudes and expectations is collected.

RESULTS

Initial research provided specific criteria and guidelines concerning the cognitive load of screen/remote interactions, the number of actions necessary to enter one character, the meaning behind symbols used, interface adaptations depending on the input end-goals, interactive ambiguities to be avoided and the application of design standards to the unique usage context of iTV. Retrospective user research is ongoing and a final compilation of data is expected to have significant effects on the initial research conclusions.



YU CENTRIK is a highly-specialized Canadian interface usability firm. Since its founding in 2003, the organization has grown to serve a wide array of large businesses through the analysis, design, and testing of interactive products.

AUTHORS

Jay Vidyarthi has a Degree in cognitive psychology from McGill University in Montreal, where he studied the human psychological tendencies which give rise to modern design principles. He has explored various avenues of human-computer interaction including academic research, web usability and user interface design.

Pierre-Alexandre Lapointe has a Master's degree in Ergonomics from École Polytechnique de Montréal. For the past seven years, he has been analysing, designing and evaluating interactive products as a UX Director at Yu Centrik.

Open Access IPTV Framework

Wei-Yun Yau

wyyau@i2r.a-star.edu.sg

Institute for Infocomm Research

Fon-Lin Lai

flai@i2r.a-star.edu.sg

Institute for Infocomm Research

Wei-Yun Yau received his PhD degree in computer vision (1999) from the Nanyang Technological University. He is the recipient of the TEC Innovator Award 2002, the Tan Kah Kee Young Inventors' Award 2003 (Merit), Standards Council Merit Award 2005, IES Prestigious Engineering Achievement Awards 2006 and the Standards Council Distinguished Award 2007. Currently, he is a Programme Manager, leading the research in Next Generation Content for IPTV.

Fon-Lin Lai received his BSc (Hons, 1st Class) in computer science (1994) from the University of Kent at Canterbury. Currently, he is the Assistant Programme Manager for the Next Generation Content for IPTV programme, and leading the implementation effort for the programme's IPTV testbed.

The **Institute for Infocomm Research (I²R)** is a member of the **Agency for Science, Technology and Research (A*STAR)** family. Established in 2002, our mission is to be the globally preferred source of innovations in 'Interactive Secured Information, Content and Services Anytime Anywhere' through research by passionate people dedicated to Singapore's economic success. I²R performs R&D in information, communications and media (ICM) technologies to develop holistic solutions across the ICM value chain. Our research capabilities are in information technology and science, wireless and optical communications, and interactive digital media. We seek to be the infocomm and media value creator that keeps Singapore ahead.

Executive Summary

The Singapore government announced an initiative to deploy a nationwide Next Generation National Broadband Network (NGNBN) which will offer pervasive ultra high speed symmetric connectivity of 1 Gbps or more by 2015, with initial provisioning of 100 Mbps from as early as 2010. To keep the network attractively priced, the Singapore government co-funds the network's deployment and also demarcated the operators into three legally distinct layers: passive infrastructure operator (NetCo), wholesale operator (OpCo) and Retail Service Providers (RSP) as shown in Figure 1. The goal for such separation is to achieve open access to the NGNBN. As video streaming is very bandwidth intensive, thus we believe the layered architecture of the NGNBN will have profound impact on the TV and video service providers over the internet protocol in Singapore, which we loosely denote as IPTV. To support multiple IPTV RSPs, an IPTV open platform is necessary. However, there is lack of dominant international standards and thus, the next best alternative is to track the ITU-T IPTV Global Standards Initiative effort which is still on-going (www.itu.int/ITU-T/gsi/iptv/). An IPTV platform operator that interfaces to the RSPs would be a better choice to start with as this permits easy modification when the standard is established. As such, the IPTV system can be rolled out while giving room for modification to the deployed system to be in conformance to the ITU-T international standards when it is eventually approved and keeping the amount of changes needed by the RSPs to a minimum through an IPTV service API. However, such a model has to be balanced with ample rooms for differentiation among the RSPs so that the open platform will promote instead of hinder innovation. The common differentiation factors are content choice and diversity, cost and level of customer service. In such an open access framework, our proposition is that it will promote the increase in content choice and diversity to cater to the varied preference of customers. With lower entry cost and nationwide access, the breakeven point for content producers is lower and thus encourages content tailored to specific customer group to be produced viably (the long tail). Other possible differentiation factors include user interface design, ease of use and degree of

personalization. Good user interface requires innovation to the Electronic Programme Guide and metadata to support recommendation instead of page-by-page search. Interactivity would also be build-in to encourage viewers' active participation instead of just passive watching. Similarly, easy and seamless operation that allows the user to enjoy his favourite content at anywhere, anytime and on any device will increase the level of user satisfaction. This necessitates heterogeneous device scalability in content streaming and conditional access that does not tie content to machines. Personalization technology allows the user to get the relevant information in the way he wants it, such as in his preferred language, at the relevant location and in the appropriate context. We believe such differentiations are just the beginning in the exciting interactive TV space. There will be other innovations in the future that will create differentiation and to attract customers since there is no monopoly and that anyone is able to introduce contents and/or services if there is business case. Thus, it will be an exciting nationwide test-bed for IPTV.

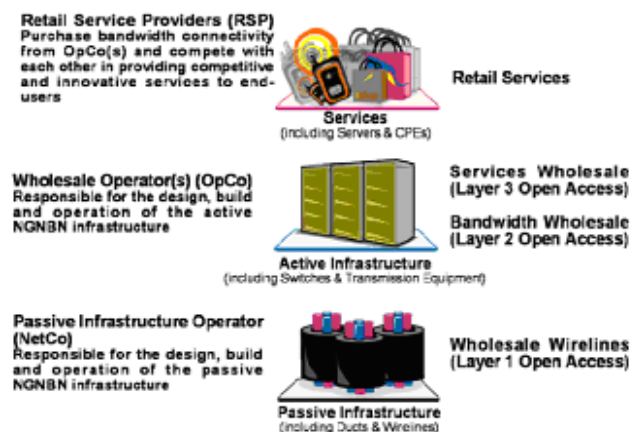


Figure 1. The Next Generation National Broadband Network Three Main Layers (courtesy of Infocomm Development Authority, Singapore)

Publishing Web Content to CE Devices

Davy Van Deursen, Frank Gielen
([davy.vandeursen,frank.gielen]@siruna.com)



1 Company & Author Biography

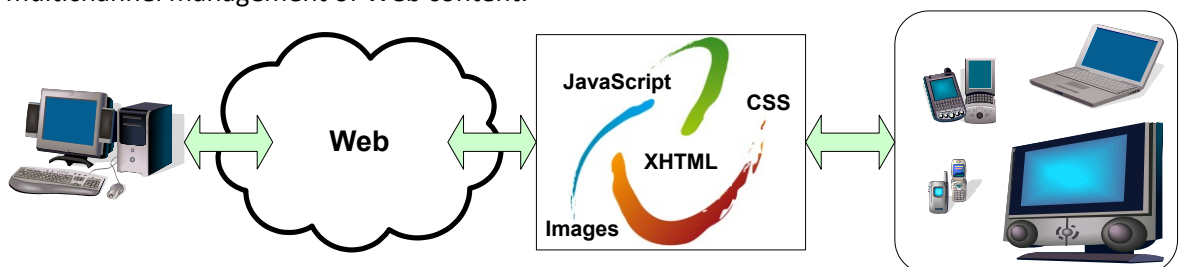
Davy Van Deursen received the master degree in computer science from [Ghent University](#), Belgium, in 2005. He joined the [Multimedia Lab](#) of Ghent University in 2005 where he is currently working toward the PhD degree. His research interests and areas of publication include video coding, media content adaptation, media content presentation, Semantic Web technologies, W3C standardization, and mobile Web publishing. Further, he is also involved in [Siruna](#), which is a spin-off of [IBBT](#) and Ghent University. Siruna is a mobile Internet technology provider, offering a platform that serves as a thin client mobile application gateway. In 2008, the company made its technology available to the open source community, implying that the platform is freely available for non-commercial purposes.

2 Web Access on CE Devices through the Siruna platform

In 2006, the [Consumers Electronics Association \(CEA\)](#) standardized the Web4CE framework (a.k.a. [CEA-2014](#)). Within this framework, they defined CE-HTML, which is a collection of technologies (XHTML, CSS TV profile, AJAX, etc.) that enable the creation of Web content for Consumer Electronics (CE) devices (e.g., TV). Several restrictions and extensions are defined on top of these technologies. Examples of extensions are support for remote control keys and media integration. In this paper, we demonstrate how the Siruna platform can be used to customize existing and new Web content for CE devices, according to the CE-HTML specification.

Siruna's primary target was to provide a software platform that customizes existing desktop Web content in real time for usage on mobile devices such as GSM's, PDA's, and smartphones. First, device characteristics and the mobile browser capabilities are detected. Based on the detected capabilities, a combination of generic and Web site specific transformation rules are applied to the input Web content. The result is a customized Web site perfectly suited for the requesting mobile device and compliant to the mobile profiles of existing Web standards.

Similar to mobile devices, bringing the Web to CE devices requires a number of changes for existing Web content. On a TV for instance, font sizes are typically higher, navigation between different links and buttons is different, the layout may be different, etc. Note that similar constraints are present on mobile devices (i.e., navigation and layout). Hence, an obvious extension for the Siruna platform is to publish existing Web content to CE devices. XHTML, JavaScript, and CSS transformations specific for CE-HTML can be easily implemented within our transformation engine. Therefore, in this presentation, we provide an overview of the necessary CE-HTML specific transformations and demonstrate how Siruna's technology enables the true multichannel management of Web content.



Talk to your customers through idtv

1 INTRODUCTION

Brands willing to communicate with their customers were facing expensive development and a long time-to-market. As a solution for this challenge, Paratel has developed an end-to-end platform where brands can communicate in a less expensive way with a short time to market.



Next to this, Paratel works together with major Belgian Broadcast companies (i.e. VMMa) to create visibility for the brands. Therefore we've created an idtv environment where the brand can buy advertising space.

For keeping costs and timings under control, Paratel has developed a set of templates that can be used by the brand. These templates are already validated on the several distributor platforms (Telenet and Belgacom).

The setup of the Paratel templates will be done through a web interface. In this way, the brand will have complete control over the message he wants to bring to the customer.



2 COMPANY & AUTHOR BIO

Paratel is 'the' one stop shop in multimedia: 'IVR, mobile applications, web, teletext and iDTV under one roof'. We offer our clients multimedia platforms for interactive communication with their end users.

Our USP's:

- Pioneer in Belgium (launch first sms and iDTV application)
- One stop multimedia shop
- In-house development
- In-house hi-tech platforms
- In-house knowledge and experience

Paul Meyers is a digital architect who has worked in the gaming industry before stepping into the idtv world. Paul has worked on several idtv projects for major brands like MTV/Nickelodeon, VTM, RTL/TVI, Belgacom, Puma and more.

User study on the impact of personalization, related content and individual preferences over interactive TV ads

We hope to inspire attendees to our presentation in creating new interactive content for ITV. The highlights of our talk include:

- A summary of the main factors explored in order to explain user engagement with ITV, as well as insights into the questions still open in this area.
- A new view on how to interpret the role of two main factors:
 - Personalization: We have analyzed the role of personalization in advertisement, studying various ways in which the content can be personalized.
 - Personal Differences: Our research has explored the impact of two particularly distinct user profiles, not based on socio-demographic characteristics but rather on their motivations for interacting with TV content.
- A look at how the evolution of our prototypes can contribute to future services for ITV. In our work we have developed innovative prototypes (Figure 1) at an initial stage, and have foreseen how these can grow and evolve to become more compelling.
- A look at our early framework for interaction with TV. Our results allowed us to answer some early queries and to present a framework draft that will provide practical guidance to industry players in the creation of new interactive content and services for ITV.



Figure 1 - Example of an offered service in one of the presented prototypes

Bio of the authors:

Pablo Rebaque Rivas obtained his degree on Psychology in 2006, from University of Barcelona, Spain. He joined Barcelona Media Innovation Centre (Barcelona, Spain), in January 2008, and has since collaborated in several research projects concerning IPTV, Internet on mobile phones and social networks.

Dr. Raquel Navarro-Prieto is currently the head of the User Interaction Lab at Barcelona Media Innovation Centre where she leads multiple research projects in the area of Human Computer Interaction: 10 national and European funded projects as well as several projects with companies such as HP, Alcatel-Lucent, Havas Media Group and Media Pro among others. In addition, she currently lectures at Universitat Pompeu Fabra (Barcelona). Previously, she coordinated all research work at the Interaction Lab of Universitat Oberta de Catalunya, Barcelona, for two years; worked 3 years at Motorola Research Lab (UK) and has industrial experience from Apple Inc. (USA), and HP (Spain). Raquel obtained her PhD on Cognitive Psychology. In addition to writing numerous papers, she has been part of and chaired several program committees for international conferences and has worked as advisor for the EU Commission.

Jorge Hernández Pablo obtained his Telecommunication Engineering degree from University of Zaragoza (Spain) in 2004. He began working for Telefónica I+D in 2003 where he has been involved in several European research projects dealing with wireless and optical networks (PULSERS I and II, Tbones). In 2007 he moved to the Advertising Technologies division and started collaborating in the i3media project, aiming on new media content technologies.

Short companies bio:

Barcelona Media Innovation Centre is a technology centre dedicated to applied research on communications and the media, and to the transfer of the resulting knowledge and technologies to the industry.

Telefónica I+D (TID) is the innovation division of *Telefónica Group*. It is the largest private R+D centre in Spain in terms of activity and resources, and it is the most active company in Europe for European research projects in the ICT (Information and Communication Technology) sector.

Broadband Internet TV For Insurance Companies

ABOUT OPTIVA MEDIA

Optiva Media is a company fully focused on Digital TV and Media projects. An established and market leader company deploying Digital TV, video and communications solutions for its customers, through a flexible, highly committed specialized services model.

ABOUT THE AUTHOR

Oscar León - Solutions Development Director

Computing background professional, with more than 15 years of experience in IT consulting and integration projects for big international consulting firms, he is focused on the technology, development and successful delivery of Optiva Media software development projects, as well as its products research and development.



PRESENTATION SUMMARY

Groupama is a European top insurance company. One of its more visionary projects has been the launch of an Internet TV service for its customers. The first TV of an insurance company in Spain. Around 2,000,000 customers, 6,000 employees and 900 mediators have access to its contents.

Groupama Seguros TV uses WebTV and video portal modules from Optiva Broadband TV. The platform, based on Adobe Flash Media technology, also allows Groupama live feeds insertions from conference rooms and mobile units, supports multiple video formats in standard and high definition, and provides functions to conduct advertising campaigns and video contests.

Optiva Media R&D activities have been key for the company product development. Optiva Broadband TV platform is continuously being improved through R&D projects. The latest platform evolution is including the automatic generation of video metadata, and ontology techniques allowing videos catalogue semantic searches.



Authors:

- *Gerrit Van den Breede*, consultant for aids for people with visual impairment in various organisations in Flanders since 1986. Training: electronics & IT, technical university level.
- *Ward De Bruecker*, consultant for aids for people with a hearing and/or cognitive impairment. Training: Masters in Psychology

Organisation: KOC – Kenniscentrum Hulpmiddelen (aid skills centre) (www.koc.be) is a service that forms part of the VAPH (Flemish Agency for the Handicapped). The KOC:

- ❖ Maintains an online database of aids with photos, moving images and prices
- ❖ Distributes information to the public via publications such as information brochures and the quarterly Infovisie Magazine (visual impairment)
- ❖ Supports the provision of advice on aids
- ❖ Advises on policy.

Accessible digital television for the visually and hearing-impaired in Flanders

Belgacom TV (www.belgacom.tv) and Telenet Digital TV (www.telenet.be) are the two biggest digital television providers in Flanders. They represent the biggest group of customers and they are the only ones to offer interactive digital television. Their digital and interactive service has been assessed by the KOC with regard to accessibility both for the blind and the visually impaired and for the deaf and the hard of hearing.

Watching digital television

Digital TV programmes are neither more nor less accessible to the blind and the visually impaired than traditional analogue television.

The visually impaired can use a bigger television set or sit closer to the screen.

The blind are assisted by a sighted person who provides an oral explanation of the images.

The KOC is investigating ways for the blind to watch television independently.

The deaf and the hard of hearing encounter difficulties with watching television because the content of TV programmes relies on speech. A TV programme becomes accessible to them when the spoken text is shown in a legible form using subtitles (teletext).

The KOC is investigating ways for people with hearing impairment to watch accessible television.

Using digital television

- Electronic programme guide (EPG)

An EPG gives a detailed summary of today's and future TV programmes on the various channels. The operation, the user interface and the language used are the elements that have an impact on the accessibility of EPGs. The KOC is investigating this.

- Video-on-demand and recording programmes with a digital video recorder (DVR) are options that the KOC is investigating with regard to their practicality.

Bringing added value to iTV users?

IP-based offers are becoming more and more popular among television users. While offers for time-shifted TV and video on demand services are offered by a variety of providers, there are only a limited number of additional services on the market, which benefit from IP technology and grant added value to the user.

In order to fill in this gap, some providers have started ambitious projects with different emphases regarding the content offered. Two of these projects are being carefully examined in a joint research study by facit digital and Alcatel-Lucent.

Alcatel-Lucent has developed a content management system, which serves as a middle-ware and facilitates to implement editorial content on iTV platforms. In the case of this study, content of a local newspaper as well as a branded entertainment channel (automotive manufacturer), are being displayed as services of Germany's biggest iTV provider.

The research interest is focusing on both, user experience aspects like usability or utility as well as users' expectations and general acceptance and of such services. Furthermore the possibilities of integration of advertising will be assessed. Therefore partly structured in-depth interviews will be conducted (partly in-home) with members of the target audience, which have access to the pilot application.

The results of the study will be available by mid may 2009 and will be presented at the conference in Leuven. They will deliver important insights about the perceived value of such services, will bring advice on how to develop user interfaces and will serve as a guideline for iTV providers, who are looking for new opportunities to broaden their range of offers to maintain successful on the market.

The authors

- **Mirja Bächle** (facit digital) holds a MA in cultural sciences and communications from the University of Munich (LMU) and has over 4 years expertise in the field of user experience, focusing on iTV applications and websites mainly.
- **Kathrin Damian** (Alcatel-Lucent) holds a MSc from the University Mittweida and has worked in the field of interactive TV since the year 2000.

The company

facit digital is a research and consulting company for digital media, based in Munich, Germany (facit-digital.com)

- Our goal is to use empirical based user research for the development and optimization of digital communication channels.
- We help our clients to optimize user interfaces and customer touch points
- We work with national and international clients from a range of business sectors, e.g. automotive, finance, entertainment and media.
- We always focus on the user's perspective.
- We are one of Germany's leading companies to provide services in user experience/usability research.

The project is realized in cooperation with **Alcatel-Lucent** (alcatel-lucent.com)

DOCTORAL CONSORTIUM

Ginga-NCL: Relating Imperative, Declarative and Media Objects

Marcio Ferreira Moreno
Departamento de Informática – PUC-Rio
Rua Marquês de São Vicente, 255
Rio de Janeiro – 22453-900 – Brazil
mfmoreno@inf.puc-rio.br

ABSTRACT

The process of developing a declarative middleware for interactive digital TV systems presents significant challenges. The main goal of this PhD thesis is to propose and to develop an NCL presentation environment for interactive digital TV systems. As one of its by-product, a declarative middleware named Ginga-NCL was developed, becoming the reference implementation of the Brazilian Terrestrial Digital TV System. This paper focuses on how NCL relationships among “distributed” applications are guaranteed by Ginga-NCL.

Categories and Subject Descriptors

I.7.2 [Document Preparation]: Languages and systems, markup languages, multimedia, hypermedia, standards. D.3.2 [Language Classifications]: Specialized application languages

General Terms

Design, Standardization, Languages.

Keywords

Ginga-NCL, NCL, DTV, Middleware, Synchronism, SBTVD.

1. INTRODUCTION

Interactive Digital TV (iDTV) applications can be understood as hypermedia documents in which related media objects of different types compose the scenes presented in receiver devices. NCL (Nested Context Language) [1], an XML based language, has become the Brazilian solution for its iDTV standard not only due to its power for easily defining spatial and temporal relationships among media objects, including viewer interactions, but also due to its facility for defining structured distributed applications across multiple exhibition devices, its support to content and presentation adaptations, its provisioning for live iDTV application producing, and its support to content and structure reuse.

The main goal of this PhD thesis is to propose and to develop an NCL presentation environment for iDTV systems. A declarative middleware named Ginga-NCL was developed as one of its results, becoming the reference implementation of the Brazilian Terrestrial Digital TV System¹. The Ginga-NCL integration in IPTV and DTV systems has brought some new solutions to

several issues regarding the support to iDTV applications, as discussed in Section 3.

An NCL document (a Ginga-NCL application) only defines how media objects are structured and related, in time and space. As a glue language, NCL does not restrict or prescribe any media-object content type. In this sense, we can have the usual media objects (text, image, video, audio, etc), imperative objects (objects with imperative code content) and declarative objects (objects with declarative code content), as NCL media objects. The media objects supported depends on the media players that are integrated in the NCL presentation engine.

Although Ginga-NCL reference implementation includes SMIL (Synchronized Multimedia Integration Language) support, the Brazilian Terrestrial Digital TV System (SBTVD) requires only the support to XHTML and NCL declarative objects. As for imperative media objects, SBTVD requires support to Lua [2] and Java (Xlet) [3] in its reference implementation. As a consequence, it is possible to create NCL applications that can use other applications (as NCL media objects) specified using the aforementioned declarative and imperative languages as well as to specify synchronization relationships among these applications.

The current research of this PhD thesis focus on both imperative and declarative object types used in NCL applications: how they can be defined, how they can be related, and how to specify the expected behavior of imperative and declarative engines through the Ginga-NCL architecture.

Considering the NCL support for multiple devices, distributed application processing allows multi-viewer interactions coming from different devices, which will allow several new interaction experiences. Another work in progress is how relationships among these “distributed” applications will be guaranteed. The multiple device support complements this thesis directions and it has been done in association with another work in progress [4].

The remainder of this paper is organized as follows. Section 2 discusses some related work. Section 3 presents the Ginga-NCL architecture and the thesis’s current contributions. Section 4 discusses future directions concerning imperative and declarative objects. Section 5 is reserved for final remarks.

2. RELATED WORK

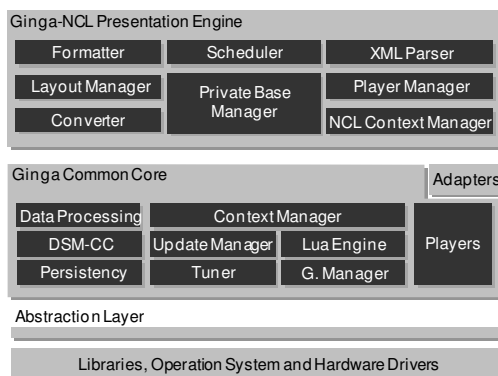
The iDTV applications can be partitioned in a set of declarative applications and a set of imperative applications [3]. A declarative application is an application whose initial entity is of

¹ An open source reference implementation of Ginga-NCL is available under the GPLv2 license: www.gingancncl.org.br/index_en.html

Most terrestrial DTV systems offer support for both application paradigms. Generally, the imperative environment of these systems is based on the use of a Java virtual machine and the definition of generic APIs that provide access to the iDTV receiver's typical resources and facilities. Usually, the declarative environment is based on XHTML user agents with ECMAScript and DOM support [3]. XHTML content authors can embed Xlets within their XHTML documents, or access Java constructs from an ECMAScript-to-Java bridge.

Unlike XHTML, NCL and SMIL have a stricter separation between content and structure, and they provide a non-invasive control of presentation linking and layout. SMIL allows the inclusion of media objects into a SMIL document, although it does not define the use of imperative and declarative objects, neither provides support for live editing and distributed processing.

As in all main terrestrial DTV Systems [3], the Brazilian middleware, named Ginga, supports both declarative applications (through its presentation, or declarative, environment named Ginga-NCL) and imperative applications (through its execution, or imperative, environment named Ginga-J [6]). Figure 1 depicts the modular architecture defined in Ginga-NCL reference implementation of SBTVD and how each module relates with other components of DTV receivers.



The core of Ginga-NCL Presentation Engine is the Formatter. This component is in charge of receiving and controlling multimedia applications written in NCL. Applications are delivered to the Formatter by the Ginga Common-Core subsystem. Upon receiving an application, the Formatter requests the XML Parser and Converter components to translate the NCL application to the Ginga-NCL internal data structures necessary for controlling the application presentation. From then on, the Scheduler component is started in order to orchestrate the NCL document presentation. The pre-fetching of media object's contents, the evaluation of link conditions and the scheduling of corresponding relationships actions that guide the presentation flow are some tasks performed by the Scheduler component. In addition, the Scheduler component is responsible for command the Player Manager component to instantiate an appropriate Player, according to the media content type to be exhibited in a given moment in time. Media contents are acquired through specifics protocol stacks, and can come from different communication networks. The author-specified relationships among media objects are respected during NCL presentations, no matter if their contents are pushed by broadcasting or pulled on demand.

Players are responsible for notifying the Presentation Engine about object events defined in NCL applications, that is, when a media segment (an anchor) begins and ends its presentation, when an anchor is selected, or when an object property changes its value. Presentation events can be derived from content timestamps (for example, PTS in MPEG-2 [7]), timers started with static contents, etc., depending on the media format. An important contribution of Ginga-NCL implementation is how to handle timebases of media contents multiplexed in an elementary stream. Algorithms to process multiple timebases multiplexed in DSM-CC NPT elementary stream [8] were defined, allowing relating interlaced stream contents with other NCL media objects.

Another major contribution of this thesis [9] is the introduction of Ginga-NCL support for live editing applications. A DTV application can be generated or modified on the fly, using Ginga-NCL editing commands [9]. These commands can be sent (or retrieved) by the service provider through a protocol stack or can be issued by imperative objects embedded in an NCL application.

153

The first one focuses on the private base activation and deactivation (openBase, activateBase, deactivateBase, saveBase, and closeBase commands). In a private base, NCL applications can be started, paused, resumed, stopped and removed, through well defined editing commands that compose the second subset. The third subset defines commands for updating an application on-the-fly, allowing NCL elements to be added and removed, and allowing values to be set to media object properties. Ginga-NCL editing commands are defined using an XML-based syntax notation identical to that used by the NCL document [1].

Ginga-NCL Presentation Engine supports multiple presentation devices through its Layout Manager module. This component is responsible for mapping all regions defined in an NCL application to canvas on receiver's exhibition devices. Currently, Resende Costa [4] thesis aims at defining a set of data structures focusing in multiple device use cases (for exhibition and also for distributed processing).

Moving back the attention to the Ginga Common Core, the Context Manager component is responsible for gathering platform characteristics and viewer profile into a data base used to update the NCL application's global variables defined in the NCL special media object called settings node. This information can then be used to adapt an application content or presentation. The settings node can also be used as any other NCL media object as an actor of a relationship.

The Players component of Ginga Common Core are composed by content decoder/players and procedures to obtain contents transported in any network supported by a specific protocol stack. The display graphical model defined by the receiver platform is maintained by the Graphics Manager component, which is in charge of handling operations on graphic planes (five in the case of SBTVD and ISDB), including overlay requests.

In the case of terrestrial DTV, the DSM-CC and Data Processing components offer support for acquiring data transported in DSM-CC carousels [8]. A resource identification mechanism was created in order to relieve application authors from having knowledge of the addressing scheme used to identify resources in DTV presentation environments and transport systems, which is also another important contribution of this thesis [10]. The Persistency component is in charge of every data storage management requested by applications. The Tuner component is responsible to offer an API for TV channels management. Finally, an Abstraction Layer was defined in order to hide implementation details, such as graphical device model, hardware decoder drivers, etc, to the Ginga-NCL and Ginga Common Core implementations

Each component of Ginga-NCL can be updated through the Update Manager component. Another contribution of this thesis is the component based implementation of Ginga-NCL, allowing component update and component loading during runtime.

4. FUTURE DIRECTIONS

As aforementioned, Ginga-NCL supports XHTML, SMIL and also nested NCL documents as NCL declarative objects. Ginga-NCL also supports Lua and Java Xlet as NCL imperative objects. NCL functionalities allow a simple definition of relationships among these objects such as, for example, "run a Lua function when a specific video segment is shown" or "submit an XHTML

form when an Xlet class gets instantiated". Figure 2 depicts a more detailed example. This NCL application is used in this section in order to clarify concepts and to illustrate the use and the specification of relationships among declarative, imperative and other media objects, which are the current focus of this thesis.

```
<?xml version="1.0" encoding="UTF-8"?>
<ncl id="SynchronizingCodes"
xmlns="http://www.ncl.org.br/NCL3.0/EDTVProfile">
<head>
  <regionBase device="default">
    <region id="hReg" width="20%" height="20%"
right="5%" bottom="5%"/>
  </regionBase>
  <regionBase device="systemScreen(2)">
    <region id="dReg" width="80%" height="80%"
top="10%" left="10%"/>
  </regionBase>
  <descriptorBase>
    <descriptor id="hDesc" region="hReg"/>
    <descriptor id="pDesc" region="dReg"/>
    <descriptor id="gDesc" region="dReg"/>
  </descriptorBase>
  <connectorBase>
    <causalConnector id="onBeginStart">
      <simpleCondition role="onBegin"/>
      <simpleAction role="start"/>
    </causalConnector>
    <causalConnector id="onEndStart">
      <simpleCondition role="onEnd"/>
      <simpleAction role="start"/>
    </causalConnector>
  </connectorBase>
</head>
<body>
  <port id="pHandler" component="handler"/>
  <port id="pGame" interface="phasel"
component="game"/>
  <media id="handler" src="scripts/handler.lua"
descriptor="hDesc">
    <property id="inc"/>
    <area id="createNextLevels"/>
  </media>
  <media id="promo" src="advert/promo.xhtml"
descriptor="pDesc">
    <area id="buyNextLevels"/>
  </media>
  <media id="game" src="game/game.ncl"
descriptor="gDesc">
    <area id="level1"/>
  </media>
  <link id="l1" xconnector="onBeginStart">
    <bind component="promo" interface="buyNextLevels"
role="onBegin"/>
    <bind component="handler"
interface="createNextLevels" role="start"/>
  </link>
  <link id="l1" xconnector="onEndStart">
    <bind component="game" interface="level1"
role="onEnd"/>
    <bind component="promo" role="start"/>
  </link>
</body>
</ncl>
```

Figure 2. Example of an NCL Application

Figure 2 presents the NCL structure module defining the root element, called <ncl>, and its child elements, the <head> and the <body> elements, following the terminology adopted by other W3C standards. The <head> element defines the following child

elements: `<regionBase>`, `<descriptorBase>` and `<connectorBase>`. There are two `<regionBase>` elements. Each one is associated with a particular exhibition device class² where presentation will take place. The `<descriptorBase>` element contains `<descriptor>` elements, each one referring a `<region>` element in order to define the initial exhibition area of a `<media>` element. A `<causalConnector>` element is specified as child of `<connectorBase>` element in order to define a causal relation that may be used to create causal relationships defined by `<link>` elements. In a causal relation, a condition shall be satisfied in order to trigger an action. Conditions and actions are specified using `<role>` child elements of `<causalConnector>` elements [1].

The `<body>` element includes `<port>`, `<media>`, and `<link>` child elements. A `<port>` element state from which media objects a document presentation chain must initiate (in the example, the “handler” and “game” objects). In Figure 2, `<media>` elements specify imperative (a Lua object) and declarative (an XHTML and an embedded NCL object) objects and its content location. Finally, the `<link>` element binds (through its `<bind>` elements) nodes interfaces with connector roles, defining a spatial and temporal relationship among objects.

The application starts with the presentation of a nested NCL object and a Lua object. The Lua object is presented on the receiver’s default screen (usually the TV set). The NCL object is a game and is presented on a class of exhibition devices named `systemScreen(2)`. All devices registered in this class shall run the game. There are two types of device classes: those able to run object players (active classes), and passive classes, which are only able to exhibit content processed in other devices. In SBTVD, class (1) is predefined as passive and class (2) as active.

The NCL game was designed with just one level of difficulty. If viewers playing the game in `systemScreen(2)` devices finish the first level, an XHTML object is started, also in `systemScreen(2)` devices, offering viewers a form of buying next game’s stages. If a viewer buys the next game’s stages, submitting the XHTML form, an XHTML embedded ECMAScript code starts the “buyNextLevels” interface. As a consequence, the Lua object interface, named “createNextLevels”, is started. This Lua object interface is mapped to a Lua function that uses the editing command API of Ginga-NCL. Thus, editing the NCL game on-the-fly, the Lua code creates and enables the next stages of the NCL game.

It is important to point out that the imperative, or declarative, player is responsible for defining how the interfaces of its NCL object type are integrated to an NCL object code chunk. In the example, SBTVD [1] specifies how the Lua player should behave to mapping the “createNextLevels” interface into a “handler.lua” code chunk (the Lua function that uses the editing command API). Another interesting work in progress is to standardize how any imperative, or declarative, player should define this code-to-interface mapping.

² In NCL, exhibition devices can register themselves in specifics classes of a domain [1] [4].

The discussed NCL application can exemplify the challenges in guarantee the relationships among NCL media objects.

5. FINAL REMARKS

The open source reference implementation of Ginga-NCL is the main contribution of this PhD thesis. During Ginga-NCL design and implementation several issues were addressed as discussed in this paper. Although initially designed for terrestrial DTV systems, the modular implementation of Ginga-NCL allows its adaptation to other platforms. Some work is currently being done in this direction, both for IPTV and peer-to-peer TV.

6. ACKNOWLEDGMENTS

I would like to thank my advisor Professor Luiz Fernando Gomes Soares for its guidance. This work is being supported by CNPq, TeleMidia Lab and PUC-Rio.

7. REFERENCES

- [1] Digital Terrestrial Television - Data Coding and Transmission Specification for Digital Broadcasting - Part 2: Ginga-NCL for fixed and mobile receivers - XML application language for application coding. Available: http://www.abnt.org.br/imagens/Normalizacao_TV_Digital/ABNTNBR15606-2_2007Ing_2008.pdf
- [2] Ierusalimsky, R. Programming in Lua. Lua.org, 2008.
- [3] Morris, S., Smith-Chaigneau, A. “Interactive TV Standards”, 2005, Elsevier Inc.
- [4] Resende Costa, R. M. 2009, Synchronization Management for DTV Applications. Submitted for publication in Proceedings of the 2009 EuroITV Doctoral Consortium.
- [5] "Synchronized Multimedia Integration Language (SMIL 3.0)", W3C Recommendation 01 December 2008. Available at: <http://www.w3.org/TR/SMIL3/>
- [6] Digital Terrestrial Television - Data Coding and Transmission Specification for Digital Broadcasting - Part 4: Volume 4 – GINGA-J: Environment for the execution of procedural applications.
- [7] ISO/IEC 13818-1. Information technology – Generic coding of moving pictures and associated audio information - Part 1: Systems. ISO Standard, 2007.
- [8] ISO/IEC 13818-6. Information technology – Generic coding of moving pictures and associated audio information - Part 6: Extensions for DSM-CC. ISO Standard, 1998.
- [9] Resende Costa, R. M., Moreno, M. F., Rodrigues, R. F., and Soares, L. F. 2006. Live editing of hypermedia documents. In Proceedings of the 2006 ACM Symposium on Document Engineering (Amsterdam, The Netherlands, October 10 - 13, 2006). DocEng '06. ACM, New York, NY, 165-172. DOI= <http://doi.acm.org/10.1145/1166160.1166202>.
- [10] Moreno, M. F., Rodrigues, R. F., and Soares, L. F. 2007. A Resource Identification Mechanism for Interactive DTV Systems. In Proceedings of the Ninth IEEE international Symposium on Multimedia Workshops (December 10 - 12, 2007). ISMW. IEEE Computer Society, Washington, DC, 215-220. DOI= <http://dx.doi.org/10.1109/ISMW.2007.6>.

Synchronization Management in DTV Applications

Romualdo Monteiro de Resende Costa

Departamento de Informática – PUC-Rio

Rua Marquês de São Vicente, 255

Rio de Janeiro – 22453-900 – Brazil

romualdo@telemidia.puc-rio.br

ABSTRACT

Synchronization support has been one of the most important QoS issues in DTV environments. In order to preserve the spatio-temporal synchronization among media objects that compound a DTV application, this PhD thesis proposes several abstract data structures to guide the synchronization control during application's life cycle. The proposed structures can be used to control the applications' playout stage at content producer and application execution at receivers in many different platforms, including those with multiple devices. The specification of applications where media objects, including those with declarative and imperative code, can be distributed to run in several different devices, recursively, is another important focus of this thesis.

Categories and Subject Descriptors

I.7.2 [Document Preparation]: Languages and systems, markup languages, multimedia, hypermedia, standards. D.3.2 [Language Classifications]: Specialized application languages

General Terms

Design, Standardization, Languages.

Keywords

Hypermedia, temporal graph, NCL, digital TV, Ginga.

1. INTRODUCTION

Temporal and spatial synchronization management in DTV applications deals not only with predictable events¹ (like the end of a media segment presentation with known duration and known beginning time), but also with unpredictable events (like viewer interactions). Content and content-presentation adaptations, which are usually performed during runtime, are other sources of unpredictability that should be supported.

In DTV applications where unpredictability is common, synchronization management is usually based on the event-driven paradigm (also called constraint/causality paradigm). Different from timeline, the event-driven paradigm bases its synchronization support on the relative spatiotemporal positioning of events, independent from when (the absolute moment in time) the synchronization happens and even if it happens.

Synchronization specifications should use high level constructs to support the authoring process. These constructs must favor relationships among media objects that exist in the author's mind model (application logical semantic). With this focus, the use of

time-based declarative languages is favored when spatiotemporal event-driven synchronization needs to be specified.

Declarative languages emphasize the declarative description of an application rather than its decomposition into an algorithmic implementation, as it is done when using imperative languages. Such declarative descriptions generally are a high-level specification, and thus they are easier to be designed than imperative ones, which usually require a programming expert. Examples of declarative languages focusing on media object synchronization are NCL (Nested Context Language) [1], the standard language of the Brazilian Terrestrial DTV System – SBTVD-T [1] and ITU-T Consented Recommendation for IPTV services [11], and SMIL (Synchronized Multimedia Integration Language) [14], a W3C Recommendation.

Usually, DTV applications are transmitted to client receivers where application language engines must try to guarantee the author's (the programmer's) descriptions. Unlike authoring constructs, presentation data structures should be closer to the execution engine and should offer low-level primitives in order to make easy the presentation scheduling. Furthermore, some support should be offered to DTV application control. As an example, it is desirable that viewers could explicitly pause a DTV application and then resume it at some later time. It is also desirable that an application could be started at any moment in time. Moreover, it would be desirable to support a viewer that changes the TV channel, starts another application in the new channel, but then regrets and returns to the previous channel, resuming the application and inheriting all actions previously triggered.

In order to assure a synchronized presentation, application data transmissions, from servers to receivers, should also be management, maintaining the minimal needed QoS.

Although authoring, presentation, and transmission goals are the same (to guarantee the synchronized presentation of media objects), these different phases of an application life cycle normally require different data structures, computed from the application specification, to support client and server sides of a DTV system [8].

In agreement with the previous paragraph, this PhD thesis proposes the use of different data structures. However, they are derived from a unique parent data structure called HTG (Hypermedia Temporal Graph) the main focus of this PhD thesis. In this paper, the use of these data structures in multiple device platforms is emphasized, since it is the current work in progress. The proposed approaches were partially put into practice in the current open source reference implementation of the Ginga (the Brazilian Terrestrial DTV) middleware [1].

¹ In this paper, an *event* denotes any occurrence in time with finite or infinitesimal duration

The paper is organized as follows. Section 2 presents some related work. Section 3 discusses the proposed structures and introduces the use of multiple devices. The application flow control and the thesis's current contribution are then discussed. Section 4 is reserved for final remarks and future directions.

2. RELATED WORK

XHTML-based languages, such as those used in BML [2], ACAP-X [3] and DVB-HTML [10] allow a declarative description of relationships involving unpredictable events; in fact, only viewer interactions. Synchronization in its broad sense can only be achieved using imperative coding, commonly written in ECMAScript [9]. In applications specified using the aforementioned languages, DOM events [13] are responsible for triggering unpredictable events. As far as we know, no one of these middleware implementations have a data structure similar to HTG to guide its synchronization tasks. Therefore, no one offers the aforementioned facilities. Indeed, it would be very difficult, if not impossible, to infer a data structure like HTG when imperative languages (in the case ECMAScript) are responsible for adaptations and spatiotemporal synchronization.

Several players are available for traditional Web-based multimedia applications [4][5]. Their data structures are fair enough for playing the majority of applications. However, DTV broadcasting has some specific features that must be taken into account. The fine-grained presentation control requested by TV channel tunings is one of them. The support for live content generation and synchronization is another. Other requirements include the fact that applications may run in receivers with limited resources. The thesis's structures for synchronization management were proposed in agreement with all these requirements [8].

The use of exhibition devices other than the traditional television commanded by a remote control, can improve viewer experience in different ways [6][7], some of them implemented by Ambulant player and Annotator [5][6]. Among other features, this player allows the exhibition of optional content in available devices, extending the application control; allows including viewer's annotations during a presentation; and allows transferring a presentation to mobile devices, without presentation discontinuity, even when the viewers are moving out (session mobility).

Non-monolithic rendering [6], with multiple devices being simultaneously used to render the application content and with multimodal interaction devices available, is one of the current research issues being addressed in this thesis. The solution proposed uses NCL as a glue language. As a glue language, NCL does not restrict or prescribe any media-object content type. In this sense, we can have the usual media objects (text, image, video, audio, etc), imperative objects (objects with imperative code content) and declarative objects (objects with declarative code content), as NCL media objects. As an example, the Ginga middleware reference implementation provides support to SMIL, X-HTML, embedded NCL, Lua and Java objects. As a consequence, it is possible to create NCL applications that can use other applications (as NCL media objects) specified using the aforementioned declarative and imperative languages as well as to specify relationships among these applications [12].

The set of NCL objects can be distributed to be processes in several different devices, under a master control. Nested NCL

objects received by devices can also be again distributed, in a recurrent process, as will be discussed in the next section.

3. THESIS WORK OVERVIEW

3.1 Synchronization Management

This research has focused on the development of several abstract data structures to support synchronization management in DTV systems. The main structure is a directed time graph model called Hypermedia Temporal Graph (HTG), which represents relationships among events in an application. When used to represent NCL applications, three types of events are recognized: presentation event, corresponding to playing a content anchor (whole media-object content, or part of this content); selection event, corresponding to a viewer interaction (selection of a content anchor); and attribution event, corresponding to setting a value to a media-object's property (variable). Each event defines a state machine, as show in Figure 1, which should be maintained by the receiver user agent. An event can be in the sleeping, occurring or paused state, and change its state upon receiving actions: start, stop, pause, resume and abort.

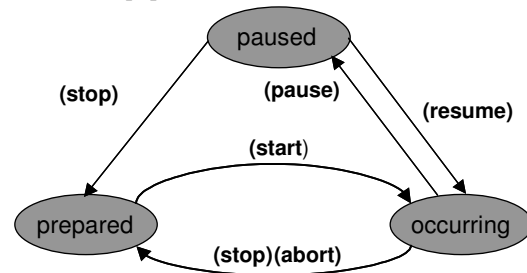


Figure 1. Event State Machine.

HTG is composed by vertices, which represent actions (for state changes) performed on event state machines, and directed edges that represent relationships among actions. An edge is labeled by a condition that must be satisfied in order to trigger the action specified in the edge's output vertice.

HTG defines simple and compound conditions. A simple condition is defined by a temporal interval that must be waited before firing the edge traverse (and so the action defined in the edge's output vertice), or by a variable that must be evaluated in relation to a desirable value, or still by external actions, such as viewer interactions. Compound conditions are defined through logical operators (or, and, not) binding two or more conditions.

After defining an application starting point, HTG can be used to derive other structures related with the synchronization management. Table 1 summarizes these structures (plans) in both client and server sides. How these plans are built and their functional details can be found in reference [8].

Table 1. Plans to support DTV synchronization management

Server Side	Client Side
Pushed-Data (Carousel) Plan	Presentation Plan
	Player-load Plan
	Pre-fetching Plan
	QoS Plan

When applications contain only predictable events, HTG edges are labeled only by temporal intervals. From the document starting time, the graph traverse identifies every action that must be applied to media players. These actions can have their moment

in time computed taking into account the time intervals required to satisfy conditions, from the HTG entry point to the corresponding action vertices. This set of actions and corresponding moments in time compose the presentation plan.

The same procedure aforementioned can be used to compute actions and their corresponding moments in time for all predictable events, from an application starting point to an unpredictable event; and from each unpredictable event to the next unpredictable event in the graph traverse. In this last case, the computed moments in time will be relative to the moment in time that the starting unpredictable event of the traverse path happens. During an application presentation, as soon as an unpredictable event time is known, the presentation plan is updated changing all moments in time relative to this event to be now relative to document starting time.

Presentation plan preserves the past and predict possible future time moments of events of an application. Using this presentation-plan feature, applications can be paused and later resumed, or still started at different runtime positions, preserving all interactions applied before the interruption.

Since the considered devices usually have resource limitations, their media players should only be instantiated when necessary. (Ginga reference implementation [1][12] is component based and allows dynamic linking). They normally cannot stay instantiated after being used, waiting for a next possible utilization. On the other hand, the time need to instantiate a media player can introduce a delay long enough to cause loss of synchronizations. A player-load plan can be used to support player instantiation management and updating, avoiding undesirable delays during the application execution.

Player-load plan is computed from the presentation plan, disregarding events other than presentation and the transitions other than the start and resume. The plan construction must take into account the delay for each specific player and platform.

3.2 Multiple Device Support

Two types of device classes were defined during the research in Ginga-NCL middleware, the presentation engine responsible for playing NCL applications: those with devices able to run object players (known as active classes), and passive classes, whose devices are only requested to exhibit content processed in other devices (registered in an active class).

Many device classes can be created. Each one must specify parameter values for the exhibition output. These parameters include the screen size, the screen graphic size and the audio type supported (mono, stereo etc). Active classes must also specify the media players they supported. Devices associated with a specific class must support all its defined parameter values. A device can be registered in more than one class, of both types.

NCL objects (including media, imperative and declarative objects) are distributed among the available devices as defined by the application authors. NCL Layout module defines the *regionBase* element whose *device* attribute allows specifying to which class an object must be delivered. By Ginga-NCL default, when only one device is addressed by an application, the *device* attribute do not need to be explicitly declared. Also by default, the class index values "1" and "2" are reserved to passive and active classes, respectively.

A device in an active class can receive an imperative or declarative object to be presented. If the object code language allows, the object's content processing can also be distributed, and so on. For example, if a device in an active class receives an NCL object with NCL code (that is, an NCL object nested in the NCL application), the processing of this object must be identical to any NCL application processing. Therefore, objects embedded in this NCL object can also be distributed in a recurrent process. In order to prevent loops, a device can not receive contents originated from its distribution process. Furthermore, devices in the same class cannot receive content from more than one device (parent device) simultaneously. These rules defines a tree structure processing distribution and will avoid several undesirable behaviors, like content presentation superposition (zIndex and graphic plan management), confusing navigation procedure (specially remote control key navigation), etc.

In Ginga-NCL the tree structure also defines to which device an interaction input must be addressed. When an NCL application is started, all input coming from any registered device must be handled by the device that runs the NCL application (the root device). That is, all devices act as input device of the root device. When a device registered in an active class receives a content to be presented and also the focus control, it gains the control of its input devices and all input devices of its descendent classes. When the object presentation finishes or when the focus control is explicitly given back, the input devices it controls are passed to the control of the parent device in the tree structure, and so forth. Objects presented in passive classes are always under control of the parent device, in all aspects.

Figure 2 illustrates a user case. In the scenario we have a TV set receiver; two headphones registered in a passive class (1); two PDAs registered in an active class (2); and a notebook, registered in an active class (3). The notebook defines another device domain, with a sound player registered in the passive class (4).

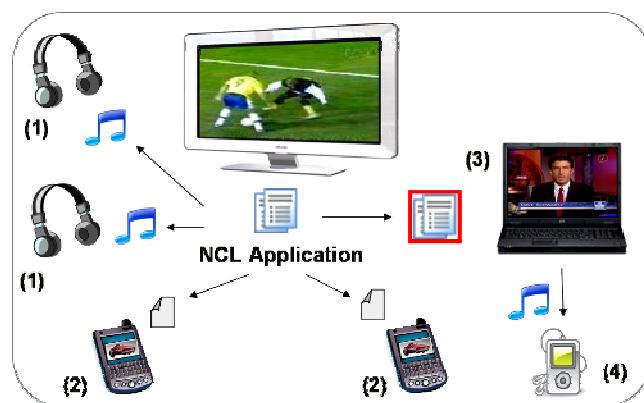


Figure 2. Example of DTV Multiple Devices Scenario.

The NCL application starts with the soccer video being presented in the TV set. During the match, audio tracks coming from the soccer field open microphone are captured and they compose the content of an audio object. This object is sent to exhibition in the passive device class (1). As a consequence, both headphones decode the audio stream sent and controlled by the TV set, simultaneously.

During the game an interactive advertisement appears. In order to avoid annoying viewers with additional information, the

advertisement is an HTML document to be played in the device class (2). As soon as this HTML object is started in class (2), each PDA instantiates an HTML user agent for its own instance of the object. The focus is also placed in these HTML object instances, so that a device in class (2) can navigate through the HTML document without interfering with what is being exhibited in the TV set or in the other PDA. For example, additional information about the product being presented can be obtained and a purchase can be done.

Instead of an HTML document, the aforementioned advertisement could be composed of a set of synchronized media objects. In this case, each PDA instantiates a specific player for each object instance received. The presentation plan, in the root device, must be used to guide the objects distribution to each PDA, where the player-load plan must be calculated to be used to support player instantiation, avoiding undesirable delays.

Also during the match, important news comes in the form of an NCL application. This application is nested in the original NCL application as one of its object. When the time to exhibit this application arrives, the NCL object is sent to the (active) device class (3). The news is then presented in the notebook under its control. During the news presentation, the NCL player in the notebook can send objects to other device classes in its own domain. For example, an audio can be played in device class (4).

In www.telemidia.puc-rio.br/~romualdo, the NCL specification of this application and the specification of the NCL object (news example) can be found.

4. FINAL REMARKS AND FUTURE DIRECTIONS

This paper presents some results and contributions of this PhD thesis and also some work in progress. Work done includes an enhanced set of data structures to support synchronization management in a DTV system. The proposed data structures preserve all temporal relationships among events, including unpredictable relationships, such as viewer interactions and those requiring on-the-fly content adaptations. Moreover, they allow starting or resuming an application at any point in time.

Current synchronization management work involves enhancing the algorithms used to calculate and maintain the synchronization plans. Elastic time computation algorithms, for stretching and shrinking media object presentations, to compensate unpredictable delays are in our plans.

One of the main motivations of this work is to provide support distributed NCL applications, in which NCL objects are played in different exhibition devices. These objects can be the usual media objects or objects whose content are imperative and declarative code spans.

As regarding multiple exhibition devices, many issues still remains to be solved. Although the main entities (classes) and relationships among them, including the presentation and focus distribution, have been defined, several implementation issues must be addressed. The next main step is to define the communication protocol among devices in the same domain, and also among devices in different domains. Implementation of other services like session mobility and viewer's annotations is being also analyzed. As proofs of concept, implementations of our

proposal for Windows Mobile, iPhone OS (Operation System) and Symbian OS are under development.

5. ACKNOWLEDGMENTS

I would like to thank my advisor Professor Luiz Fernando Gomes Soares for its supervision. This work is being supported by Telemidia Labs and PUC-Rio.

6. REFERENCES

- [1] ABNT 2007. Digital Terrestrial Television Standard 06: Data Codification and Transmission Specifications for Digital Broadcasting, Part 2 – GINGA-NCL: XML Application Language for Application Coding (Brazil, November, 2007).
- [2] ARIB Standard B-24 Data Coding and Transmission Specifications for Digital Broadcasting, version 4.0, 2004.
- [3] ATSC Advanced Television Systems Committee. 2000. ATSC Data Broadcasting Standard - A/90, 2000.
- [4] Buchanan M.C., Zellweger P.T. Specifying Temporal Behavior in Hypermedia Documents. Proceedings of European Conference on Hypertext (Milan, Italy, December 1992). ECHT'92.
- [5] Bulterman D.C.A., Jansen J., Kleanthous K., Blom K., Benden D. AMBULANT: A Fast, Multi-Platform Open Source SMIL Player. In Proceedings of ACM International Conference on Multimedia (New York, USA, 2004).
- [6] Cesar, P.; Bulterman, D.C.A.; Obrenovic, Z.; Ducret, J.; Cruz-Lara, S. 2007. Non-Intrusive User Interfaces for Interactive Digital Television Experiences. In Proceedings of European Interactive TV Conference (Amsterdam, Netherlands, May). EuroiTV 2007.
- [7] Cesar, P.; Bulterman, D.C.A.; Jansen, J.; 2008. Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share and Transfer Television Content. In Proceedings of European Interactive TV Conference (Salzburg, Austria, July). EuroiTV 2008.
- [8] Costa, R.M.R.; Moreno, M.F.; Soares, L.F.G. 2008. Intermedia Synchronization Management in DTV Systems. In Proceedings of ACM Symposium on Document Engineering (São Paulo, Brazil, September). DocEng 2008.
- [9] ECMA International - 1999. ECMA – 262 – ECMAScript Language Specification. 3rd Edition.
- [10] ETSI European Telecommunication Standards Institute. 2006. ETSI TS 102 812 V1.2.2 Digital Video Broadcasting “Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.1.1”.
- [11] ITU-T Consented Recommendation H.761. Nested Context Language (NCL) and Ginga-NCL for IPTV Services. 2009.
- [12] Moreno, M.F. Ginga-NCL: Relating Imperative, Declarative and Media Objects. Doctoral Consortium of European Interactive TV Conference (Belgium, June). EuroiTV 2009.
- [13] W3C World-Wide Web Consortium. 2004. Document Object Model – DOM Level 3 Specification. W3C Recommendation
- [14] W3C World-Wide Web Consortium. 2008. Synchronized Multimedia Integration Language – SMIL 3.0 Specification, W3C Recommendation.

Designing iTV Interfaces for Preschool Children

Ana Vitória Joly
University of Brighton
Lewes Road
Brighton, BN2 4GJ, UK
+44 7716177112

A.V.B.Joly@brighton.ac.uk

ABSTRACT

The aim of this research is to investigate preschoolers' interactions with interactive television applications. The study involves the design of a prototype and empirical evaluation in order to produce guidelines and refine techniques of design and evaluation involving young children.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *User-centred design*.

General Terms

Design, Experimentation.

Keywords

Children, Interactive Television, User-centred Design.

1. INTRODUCTION

This research is focused on the design and evaluation of interactive television interfaces for preschool children.

During the past years, research reported that screen entertainment could affect young children's development and guidelines were released first recommending that children under the age of two should not watch television [24] and more recently, a British study suggested that children under three year old should have no screen exposure [21]. However, despite the extensive research about the harms screen media could cause to children's development such as obesity [19], attentional disorder [4] and aggression [9]. There are several studies that highlight the benefits of television for cognitive and social development [6], [2], [5], [10]. The fact is, today's children have increasing access to a range of electronic media and different content is specially produced to each particular age group, denying preschoolers access to this content would prevent them to experience the legitimate benefit of media.

Research indicates that, among a variety of media devices and content such as computer games, the internet and portable electronic devices, television still appears to be a secure part of

children's cultural 'diet' and it is more important than other media because of its universal accessibility to all classes, ages and types of children [7].

Children's television has always been interactive, viewers since the early ages of TV were invited to dance, sing along, make arts and crafts. In the 1950s a programme pushed the boundaries and asked children to interact by drawing on the television screen using a special kit to help the character with what he needed during his adventures; broadcast in the United States on Saturday mornings *Winky Dink and You* is considered the first interactive television (iTV) programme.

There are several definitions for interactive television. It could be strictly related to programmes that are digital broadcast and make use of a return channel to establish a dialogue between the viewers and the broadcaster. Alternatively it could include interactivity through telephone, letters or simply by asking the audience to perform activities as described above. In this research, interactive television is defined as programmes, applications and services that the user interacts with using a device connected or directly linked to the screen (e.g. remote control, mouse). In this case the user is able to interact and alter the audiovisual content being displayed while the interactivity could take place locally at the set top box or another type of receiver or via return path and it is not limited to traditional TV.

2. AIMS AND OBJECTIVES

Currently there are very few studies on interactive television for children. Chorianopoulos and Lekakos (2007) explore the characteristics of interactive TV that facilitate education and play focused on a wider and older age group [3]. Hynd (2006) examines comprehension, attention or enjoyment of young children viewing different types of interactive TV programmes compared with those of children viewing non-interactive versions of the programmes [13]. Interactive television research still focuses on children as merely testers evaluating the interactive services at the end of the design process. Research on human computer interaction identifies user involvement as crucial to gain a better understanding of user needs and goals leading to a more useable product [17]. The interaction design for children literature had also already underlined the importance to have users contributing more actively during the design process.

In research with preschoolers, a participatory design approach involving them as design partners [8] would be time consuming and probably veer the focus of the analysis of their interactions towards solutions for an specific design. On the other hand, children participating as mere testers do not elucidate as much detail of their relationship with the media. If involved earlier in the research, they are more likely to give suggestions and talk about different ways of interaction than during usability testing

sessions with high-tech prototypes [20]. As a result, the informant design approach was chosen, as it resides between the user centred design and participatory design [20].

This qualitative research in which preschoolers are informants is not focused on the characteristics or effects of the media but on empirical evidence of how 3 and 4 year old children, the youngest users allowed by medical community, interact with the television, considering them viewers not learners.

More specifically, the aims of the proposed research are: (1) analyse these interactions complexities and details in order to further the understanding of the way preschoolers interact with the television, (2) contribute with design guidelines for preschool interactive television, (3) refine methods and add to the knowledge of design and evaluation techniques involving young children. In order to achieve these aims a prototype of an electronic programme guide (EPG) was developed with participants' contributions and is being tested. Video data combined with survey results gathered during the testing sessions will be analyzed and used to identify and articulate a set of related concepts which should constitute an integrated framework that can be used to explain or predict preschoolers' interactions with iTV applications. Then guidelines and refined methods will be presented to assist on the design and evaluation of interfaces that meet the needs, capabilities and interests of preschool children.

3. METHODOLOGY

As stated previously the main aims of the study is to analyse how preschoolers interact with iTV applications to infer design guidelines, refining methods of design and evaluation during the process, therefore the research approach is qualitative in nature.

A preliminary literature review covering children and television, interaction design for children and interactive television was done in the first stage of the research to enhance theoretical sensitivity and also to inform the design of the low-tech prototype. And then children were involved in the design process with the purpose of obtaining more information about the specific age group, to refine user requirements and improve the prototype and also in order to assist on structuring the framework.

3.1 Design Methods

3.1.1 Observing Children

The first stage of data collection involving children was the observation. The objectives of observing children in the nursery setting were: Familiarization with the age group, obtain inspiration and ideas, an opportunity to know children and their behaviour and for them to feel more comfortable with the researcher during the following stages of the study.

Twelve hours were spent observing five children, and these data was collected in form of field notes. The data was coded as collected, and categorized, actions and procedures were defined and informed the first version of the framework and prototype. Different concepts such as "play mode and style" and "preferences and ownership" were related to emerging categories such as "playing alone/sharing/competing", "favourites/likes and dislikes/sense of ownership" that reflected children's actions and behaviours and indicated prototype's requirements such as flexibility for one or more users and customization.

3.1.2 Card Sorting Activities

The observation sessions did not elicit enough data to establish

the categories to be implemented in the EPG, as a result, in order to define and refine these categories it was decided to conduct card sorting activities with children.

Card sorting techniques have been used to create information architecture eliciting conceptual structures from participants in order to reflect how users view the content [16]. Card sorting tasks have been carried out with children as young as eight years of age [12]. In this research this technique was adapted to be used with 3 and 4 years old children and contribute to the design of technology for this age group.

3.1.2.1 Closed Card Sorting

The first card sorting activity was inspired by the Dimensional Change Card Sorting (DCCS) task used to determine extradimensional shifting abilities in preschool children [14]. For the prototype development, an initial set of categories were pre-determined based on children's TV channels websites, children's film and book categories. The closed card sorting was then used to check how well those categories fit children's expectations.

There were six sessions conducted with fifty six children in five different nurseries. Each child was tested individually in a session that lasted for approximately ten minutes. Materials consisted of two side-by-side shoeboxes each with a plastic sheet to display a category plus a slot through which the child posted a laminated card showing a screenshot from a video. Two predefined categories were displayed at a time and the child's task was to post the card in the box they found more appropriate. In all eleven screenshots would have been sorted.

Results from this study indicated very strong evidence that participants were not guessing while conducting the activity, therefore preschoolers are capable of categorizing audiovisual content and would then benefit from an EPG with appropriate categorization. The activity also showed that children related screenshots to its video content; hence screenshots could represent videos on the EPG interface. According with the number of screenshots posted in the expected category box, results indicated that some predefined categories such as "Super Heroes" and "Fairy Tales" were well understood while other categories like "Make and Do" and "Around the World" were not as comprehensible for preschoolers so needed to be redefined or eliminated. It was also found that, despite the increasing number of children's channels broadcasting the same content and video on demand features that disconnect the content from the channel, the number of children who recognized the channel in which videos were broadcast was significant, so children's channels were included as categories in the EPG prototype.

3.1.2.2 Open Card Sorting

The categories tested by the closed card sorting were made for children, by adults. There was a concern that child based categories of audiovisual content could be completely different from what was pre-established based on existent categorization; therefore, it was decided to conduct an open card sorting activity. Based on the Hierarchical Taxonomic Concept Test [22] the task was explained to the participants as the grouping game, in which they had to put together things that are the same type or kind. They were handed sets of cards and given some time to make groups before the next set was handed. Children were motivated to consider all screenshots while making groups, not just the separated sets. And at the end of the activity children were asked to justify their choices.

This activity did not elicit design decisions. Children usually become overwhelmed when too many options were shown and could not associate them, nor could they explain their choices. It was found that this task has to be further developed to help in the design; it may require more of participants' time and would be probably necessary to have several sessions to achieve some level of contribution to the information architecture of a system.

3.1.2.3 Match-to-Sample

The results from the closed card sorting indicate which categories are well understood by children and which ones are not as clear. In case most participants relate a screenshot to the expected category the design decision is simple, to maintain that category. However when a category is not comprehended the design decision could be either eliminate or refine the category, and to do so it is essential to identify to which other category children would relate its members. In order to confirm the closed card sorting results and find if the screenshots that were not inserted into the expected category could fit within another pre-established category instead the match-to-sample activity was developed.

Mervis and Pani (1980) study was adapted organizing twenty four screenshots into six different groups of four screenshots each [15]. The groups were made by some of the pre-defined categories. The screenshots were printed in colour on A4 paper. Children were given a sticker with a black and white screenshot and asked to put it on the group they thought it would be most appropriate. After the child had chosen a group and had pasted the sticker, s/he was given another A4 sheet with the same categories but in random order and composed of different screenshots, and another sticker with another screenshot to be pasted. There were six different screenshots to be pasted on six A4 sheets.

Following this activity categories for the EPG were refined. The results suggested that children are able to cross-classify screenshots based on videos' type (e.g., movies), format (e.g., cartoons) or genre (e.g., fairy tales) suggesting that they are not restricted to a single form of categorization. This indicates that in an EPG children could benefit from a significant overlap in categories rather than one replacing the other. Therefore, for the prototype under development it was decided to make each category broader, so it could include all or most participants' grouping choices, and overlap the categories.

3.1.3 Low-Tech Prototyping with Children

Previous data collected informed the prototype being developed but there were still a lot of issues to be clarified such as icons to be used and where to place them on the screen. At this point it was decided to ask children for direct input on the "look and feel" of the interface. Scaife and Rogers (1999) suggestions for low-tech prototyping with children, such as the use of laminated images which could be manipulated against a background, were combined with some ideas to work with younger children as design partners [11] to create a session appropriate for this age group but not as time consuming as the cooperative inquiry.

During the low-tech prototyping session preschoolers were asked for input and suggestions. Children were told we were working on a "programme finder" and needed their help. They were given a A3 paper in which a TV set was printed, then each one received the first screenshot and asked to choose one icon among three options provided that would be more appropriate to help children find the particular programme. The same procedure was carried

out for nine screenshots and then they were asked to choose one icon to help children find "help", one to close or exit the "programme finder", one to find their favourite programmes. Children were then provided with glue and crayons to create with the material chosen their own EPG.

The icons chosen by most participants were the ones used in the EPG prototype. And from this process it was also possible to have a glimpse on how children understand the interaction process, as they paste the images and draw on their low-tech prototypes they talked about what they were doing and what would happen during the interaction with their "programme finders".

3.1.4 Prototype Adjustments with Children

There was the concern that some of the icons could not be easily recognisable. The icons used on the prototype were chosen with children during previous stage of the research, but they were provided with a very limited amount of options. So it was decided to test if the 3 icons, that were most criticized by the experts in their evaluation, could be replaced by a more meaningful option.

In this session, children were explained how the prototype worked. Then they were asked to perform the same task as the experts, however they were not told the exact steps to be followed, so tips were provided if they got stuck. As soon as they accomplished the task another version of the prototype in which other options for icons were included was shown. Children were asked if they could identify the icon for a certain function, then they were shown two other options and asked which one amongst the existent and the two other options was the most appropriate icon. The optional icons could be dragged and dropped above the existent one replacing it. Following children's suggestions, one of the 3 icons tested was replaced after this session.

This activity could probably be related to a high-tech prototyping session, yet prototype adjustment with children was found more appropriate because it was decided that at this stage the only crucial input needed was to check the appropriateness of icons so children were not given enough room and structure to opinion.

3.2 Evaluation Methods

3.2.1 Expert Evaluation

Experts from academic and industry specialized in interactive technologies, children and/or interactive television were asked to analyse prototype and make suggestions for improvement. This analysis was important to improve the prototype to be shown to children with a reduced number of navigational problems.

First experts were asked to conduct a cognitive walkthrough [23], in which they performed a task and should check for each step how easy it was for a new user to accomplish the task. The walkthrough was used as a way to provide structure to explore the prototype, and then they were asked to answer some questions with their opinion about the system. These questions were based on the structured expert evaluation method [1], an analytical evaluation method designed to assess fun and usability of young children's computer games and adapted to suit an iTV application.

Thirteen experts from six different countries evaluated the prototype. The evaluations assisted on improvements on the prototype and together with the literature review also helped to stimulate thinking about properties and furnish initial ideas.

3.2.2 Testing Sessions

This is the stage the research is currently in. The EPG prototype was tested with children in two Nurseries in São Paulo, Brazil and one Nursery in London, UK. An additional session will be run in another Nursery in London.

Children were asked to interact with the prototype using the mouse and remote control, and their interactions and facial expressions were recorded. After testing the prototype participants were asked to fill in the Fun Toolkit survey [18] in which they rated with smiley faces the experience they had using the mouse and remote, if they would like to use any of the devices again to interact with the prototype and which device was the best one.

The data from the testing sessions will be analyzed and as soon as it is decided which categories best explain what happens in the study, they will be treated as concepts to generate the framework.

4. MAIN CONTRIBUTIONS

It is expected that the data collection and analysis process described will result on a set of concepts related through statements of relationship, which together constituted an integrated framework that can then be used to explain or predict how young children interact with iTV applications. These results will then be used to develop design guidelines for preschool interactive television. During this process methods are being refined to be added to the knowledge of design and evaluation techniques involving young children.

5. ACKNOWLEDGEMENTS

A special thanks to children and staff from: One World, Bright Start, Kipling Lions, Roundabout, Turnerland and Wishing Tree Nurseries (Brighton); Trenzinho Feliz and EMEI Clycie Mendes Carneiro (São Paulo), the Childcare at Lara and Barnaby Bright (London). I would also like to thank Dr Lyn Pemberton, Richard Griffiths for all their support and guidance.

6. REFERENCES

1. Baauw, E., Bekker, M.M. and Barendregt, W. A Structured Expert Evaluation Method for the Evaluation of Children's Computer Games. in *Human-Computer Interaction - INTERACT 2005*, Springer Berlin, 2005, 457-469.
2. Baydar, N., Kağıtçıbaşı, Ç., Küntay, A.C. and Gökşen, F. Effects of an educational television program on preschoolers: Variability in benefits. *Journal of Applied Developmental Psychology*.
3. Chorianopoulos, K. and Lekakos, G. Learn and play with interactive TV, ACM Press, 2007, 4.
4. Christakis, D.A., Zimmerman, F.J., DiGiuseppe, D.L. and McCarty, C.A. Early Television Exposure and Subsequent Attentional Problems in Children *Pediatrics*, 113 (4). 708-713.
5. Close, R. Television and language development in the early years: a review of the literature, National Literacy Trust, London, 2004, 42.
6. Davies, M.M. *Television is Good for Your Kids*. Hilary Shipman, London, 1989.
7. Davies, M.M. and Thornham, H. Academic Literature Review: The future of children's television programming, Ofcom, 2007, 26.
8. Druin, A. Cooperative inquiry: developing new technologies for children with children *Proceedings of the SIGCHI conference on*

Human factors in computing systems: the CHI is the limit, ACM Press, Pittsburgh, Pennsylvania, United States, 1999.

9. Eronen, L. and Vuorimaa, P. User interfaces for digital television: a navigator case study *Proceedings of the working conference on Advanced visual interfaces*, ACM, Palermo, Italy, 2000.
10. Gentzkow, M. and Shapiro, J.M. Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study. *The Quarterly Journal of Economics*, 123 (1). 279-323.
11. Guha, M.L., Druin, A., Chipman, G., Fails, J.A., Simms, S. and Farber, A. Mixing ideas: a new technique for working with young children as design partners *Proceeding of the 2004 conference on Interaction design and children: building a community*, ACM Press, Maryland, 2004.
12. Hanna, L., Risden, K., Czerwinski, M. and Alexander, K.J. The Role of Usability Research in Designing Children's Computer Products. in Druin, A. ed. *The Design of Children's Technology*, Morgan Kaufmann Publishers, San Francisco, 1999.
13. Hynd, A. Evaluating four and five-year old children's responses to interactive television programs., Murdoch University, 2006, 313.
14. Kloo, D., Perner, J., Kerschhuber, A., Dabernig, S. and Aichhorn, M. Sorting between dimensions: Conditions of cognitive flexibility in preschoolers. *Journal of Experimental Child Psychology*, doi:10.1016 (j.jecp.2007.12.003).
15. Mervis, C.B. and Pani, J.R. Acquisition of Basic Object Categories. *Cognitive Psychology*, 12. 496-522.
16. Nielsen, J. Card Sorting to Discover the Users' Model of the Information Space, 1995.
17. Preece, J., Rogers, Y. and Sharp, H. *Interaction design: beyond human-computer interaction*. John Wiley, New York, 2002.
18. Read, J.C. and MacFarlane, S. Using the fun toolkit and other survey methods to gather opinions in child computer interaction *Proceeding of the 2006 conference on Interaction design and children*, ACM Press, Tampere, Finland, 2006.
19. Robinson, T. Television Viewing and Childhood Obesity *Pediatric Clinics of North America* 48 (4). 1017 - 1025.
20. Scaife, M. and Rogers, Y. Kids as Informants. in Druin, A. ed. *The Design of Children's Technology*, Morgan Kaufmann Publishers, San Francisco, 1999.
21. Sigman, A. Visual voodoo: the biological impact of watching TV. *Biologist* 54 (1). 6.
22. Sung, Y.-T., Chang, K.-E. and Lee, M.-D. Designing multimedia games for young children's taxonomic concept development *Computers & Education*, 2008, 1037-1051.
23. Wharton, C., Rieman, J., Lewis, C. and Polson, P. The cognitive walkthrough method: a practitioner's guide in Nielsen, J. and Mack, R.L. eds. *Usability inspection methods*, John Wiley & Sons, Inc., New York, 1994 105 - 140
24. Zimmerman, F.J. and Christakis, D.A. Children's Television Viewing and Cognitive Outcomes A Longitudinal Analysis of National Data. *ARCH PEDIATR ADOLESC MED* 159. 7.

User-centered multimedia content personalization: a services composition based approach

Marcelo G. Manzato

Mathematics and Computing Institute – University of Sao Paulo
Av. Trabalhador Sancarlense, 400, PO Box 668 – 13560-970, Sao Carlos, SP – Brazil
mmanzato@icmc.usp.br

ABSTRACT

In this project, we propose a distributed architecture which supports metadata extraction by exploring interaction mechanisms among users and content. The interaction activity addressed in this work is related to peer-level annotation, where any user acts as author, being able to enrich the content by making annotations, using, for instance, pen-based devices. The exploration of interaction mechanisms mentioned here will be implemented as services compositions, as well as the provision of adaptation tasks, so that users will be able to use portable devices to access and interact with multimedia content.

Categories and Subject Descriptors

J.9 [Mobile Applications]: Multimedia applications and multimedia signal processing; I.2.13 [Artificial Intelligence]: Knowledge Management—*Knowledge personalization and customization, knowledge acquisition, knowledge modeling*

General Terms

Design

Keywords

Personalization, adaptation, interaction, services composition, peer-level annotation.

1. INTRODUCTION

One of the greatest advantages that digital television has brought to users is the capability to interact with the content [15]. This affirmative can be supported if we look at the huge advances of the web, whose main characteristic is the interactive scenario where users are able to choose alternative navigation paths, explore different pieces of information and click on related links they wonder to check. The interaction with multimedia content, in special, is obtaining extra development efforts, because of the recent availability

of web-based authoring tools, such as YouTube, Facebook, etc. Those tools have changed the user-consumer role into a user-producer role, enabling any user to create multimedia content and make it available on the web.

In parallel with the development of interactive web applications, it is becoming usual the use of portable and mobile devices to access multimedia content. This is due to the increasing processing power of those devices and because multimedia content can be manipulated in the sense of being adapted according to some particular situation. This last possibility, indeed, is issue of the content adaptation and personalization research area, whose adaptation is defined by Lum & Lau [12] as the applications' ability to choose the best version of the content to be successfully accessed by users with devices containing restricted capabilities. The personalization, in turn, is defined by Barrios et al. [4] as a particular case of multimedia adaptation where the data is adapted according to the needs and preferences of a specific user.

The multimedia adaptation and personalization, the use of portable devices, and the interaction functionalities have a relationship that is worth to consider. Firstly, users are more likely to interact with the content if they are using portable devices. Secondly, to make annotations onto multimedia content using portable devices, it is needed some kind of content adaptation, so that it can be successfully accessed and annotated by the users [12]. Lastly, as personalization services provide adapted content according to a user's preferences and needs, which may require a previous semantic knowledge about the environment in order to decide the best version of the content, the extraction of this high level information may be better accomplished if we explore some clues given by the user at the interaction time.

The interaction itself, as argued by some authors [7], can be classified in content consumption, setting selection, navigation/selection and authoring. This last category, as previously mentioned, obtains extra attention from users as it allows them to act as content providers and personalize the data, adding valuable metadata as additional information semantically related.

Usually, the authoring process can be done following two different approaches [5]: hierarchical, which provides information about specific media items with the objective to be searched or analyzed. One example could be metadata referring a movie, such as title, producer, list of actors, etc. The second authoring approach is called peer-level annotation, and can be accomplished by any person. One example mentioned by Cesar et al. [7] is the highlight of an actor's

text name using electronic ink. Two main characteristics of this level of annotation are that it does not follow a restrictive vocabulary [7], and that usually the annotation is made using portable devices, which have limited capabilities.

The need of semantic metadata to create personalization services certainly can be supported by hierarchical content authoring; however, the literature [1][18] reports that the job of annotating is time-consuming. Furthermore, producers will annotate certain characteristics that he/she subjectively thinks it is important for future applications.

Consequently, the exploration of peer-level annotation may minimize the hierarchical problems mentioned above. Firstly, the time-consuming question can be partially solved if multiple users dedicate collaborative effort to the same content. Secondly, the subjective choice of different aspects of the content to be annotated has a slighter drawback if we consider that the information extracted by the user interaction will be used for personalization services, which, in turn, will benefit the same user or set of buddies that usually shares the same preferences.

However, the use of peer-level annotation brings challenges that need further research. One of them is the fact that interaction between user and content can be done in different ways. Thus, a number of techniques must be available in order to analyze the interactive scenario and find valuable information that can be considered metadata. This set of techniques, indeed, sometimes needs to be combined with each other depending on the interaction activity executed by the user in a given period.

Another challenge is the provision of adaptation services, together with interaction functionalities, that must be considered when peer-level annotation is accomplished using portable devices. In this case, the complexity increases at the moment that the content to be accessed by the user has time constraints, such as interaction's response time or real-time applications.

Lastly, the peer-level has the characteristic of not following a restrictive vocabulary; consequently, algorithms to convert interaction-based extracted data into a representative format must be developed.

2. OBJECTIVES

As users usually enjoy to interact with the content, making personal annotations, using, for instance, pen-based devices, and maybe they wish to share those annotations with friends, this project proposes a distributed architecture with services compositions to support automatic metadata creation based on peer-level annotation at user/content interaction time. This information will, in turn, act as meaningful metadata for application scenarios, such as multimedia personalization. The main advantage of this approach is that users won't be bored as they usually do when using human-assisted tools to create content metadata, because they will be enriching the content without worrying about the true application's intention behind the interface. When necessary, the system will also be able to combine and use adaptation services to make possible the access of multimedia content using portable devices.

Considering the challenges of using peer-level annotation to extract metadata, the proposed architecture will address them as following:

- Combine different techniques (or services) that explore

particular interaction aspects and adaptation procedures, such as handwriting recognitions, highlighted object segmentation, low-level features extraction, multimedia content transcoding and transmoding, etc.

- Isolate the unrestrictive annotation vocabulary by adding a representation model for metadata extracted from the interaction. This modeled information will then be available to any application that wishes to use them for personalization tasks.

3. RELATED WORK

The generation of metadata is a time-consuming task, and hence, should not be delegated only to content providers, as more content will be available each day. In addition, some authors report that users are usually loath to do manual annotation, and also, automatic analysis doesn't reach powerful results for the applications' goals which are wanted to have [19][6]. Consequently, the insertion of consumers into the metadata authoring process is a interesting design plan, being mentioned by many publications available on the literature.

Nack & Putz (2001) [17] and Gemmell et al. (2005) [11] use additional equipment and processing capabilities to capture conceptual dependencies at specific times of video creation. If on the one hand metadata can be extracted from these supporting technologies, on the other hand the user must be able to handle all the computational environment.

Considering this drawback about the need of user intervention, some pieces of work explore the fully automated metadata extraction activity. Lots of authors have published work that proposes techniques to gather related information about the content. Venkatesh et al. (2008) [19] cite some ways to extract this information; however, most techniques has good performance only in specific video domains.

As an attempt to solve the specific video domain problem, some authors [9] have published papers that deal with media aesthetic [20], which is a study and analysis of media elements such as lighting, motion, color, and sound both by themselves and their roles in synthesizing effective productions. However, each person will interpret the content information in a different way, and thus, users which don't have the same sensori-emotional values as those defined by the technique will have the possibility to not being satisfied.

Considering the fact previously mentioned that users may have different reactions for the same content, some work tries to model the user behavior in order to achieve semantic information extracted from multimedia content. Most work [10][8] models the user activity; however, they do not study the user annotation when interacting with the content.

Following research that deals with user models, authors have focused on user-centered approaches in order to provide personalization tasks according to personal preferences. Some publications, for instance, propose techniques based on personal traits [2]; but required mapping between user's preferences and audiovisual content features is not a trivial task to be accomplished.

Visual features, in special, are used in [3] in order to improve the metadata generation in a collaborative way. However, more study is necessary in order to determine when and how collaborative authoring and automatic metadata extraction can be all combined [19].

For collaborative work, user-produced data must be avail-

able for sharing in a way to be explored by applications. Content enrichment is explored in an architecture proposed by Cesar et al. (2006) [7] which supports authoring of additional information by users, and sharing of this data among users; however, no metadata is created by exploring the additional content which is inserted by the user.

At this point, considering the state of art previously depicted, it is possible to note that the use of peer-level annotations is suitable to extract meaningful metadata from multimedia content. The main techniques' limitations, which were described in this section, and that can be further explored are: i) the lack of techniques that use collaborative work from different people to annotate content; ii) the time-consuming efforts to author metadata; iii) the content-centered approach which generalizes user's perceptions at visualization time; and iv) the domain-specific techniques which are supposed to be fully automatic.

4. METHODOLOGY

The methodology, in this text, is defined as the description of how the project could be developed. Firstly, this section depicts a set of techniques that can be used to extract metadata by exploring peer-level annotation. Lastly, an initial version of the architecture is presented, which can contribute to better understand the purposes of the work, and how they will be achieved. It is important to mention that partial results of the ideas presented here are already available on a publication [13].

4.1 Metadata Extraction

In the following subsections, we selected a set of issues that can be used to gather additional information from the user as he/she is interacting with the content.

4.1.1 Individual frame(s) or scene(s) chosen by the user to annotate

Low-level characteristics of chosen frames/scenes gathered from a video sequence may give information about content that is interesting to the user. These low-level characteristics may include histograms and color information, motion-field among successive frames, presence or absence of specific objects, etc.

Furthermore, users may restrict the access of specific scenes, once it may contain, for instance, inappropriate content for their children, or it may contain subjects that are out of scope of their preferences. Using classification techniques [14] [16], it is possible to define what kind of content will probably be denied or accepted by the user.

4.1.2 Annotation gestures and assistive segmentation

Related work about object segmentation usually has drawbacks, such as over-segmentation, when using those techniques to extract objects of interest without user's assistance. As users normally make annotations near interesting objects, some algorithm may be developed to assist the segmentation procedure in the sense of finding a relationship between the wanted object and the annotations nearby. As soon as the interesting object is extracted, the personalization task, in turn, may be benefited by object recognition and feature extraction algorithms.

When considering only the annotation activity itself, some handwritings' characteristics, such as shape, intensity and

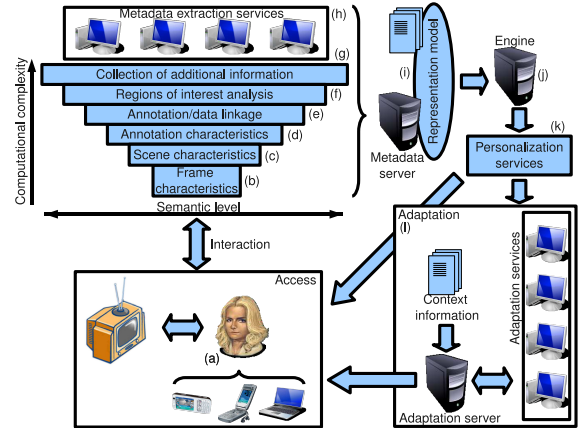


Figure 1: Proposed architecture for peer-level-based personalization.

speed, may give information about the user's sensation, feeling and level of interest about the content. For instance, circumferences around a specific area may imply that an important piece of information is present into that circle.

4.1.3 Content sharing among users, documents and interfaces

Usually, users have friends with similar tastes. Shared information gathered from a buddy may be used as additional metadata to identify the user's preferences. Discovering this relationship by comparing related documents may also inform what subject has triggered the user's attention, and hence, personalization tasks may explore this data in order to provide content according to user's tastes.

In addition, together with handwriting recognition, a number of different interfaces may be used for interaction. Voice recognizers, mobile phones, remote controls, tablet augments, etc., all of them have specific characteristics that can be analyzed in order to extract information about the user's preferences.

4.2 Initial Architecture

According to the different peer-level-based metadata extraction methods that were depicted in last subsection, we describe in this subsection an architecture that gives support to the exploration of interaction capabilities among user and content. Figure 1 presents the overall schema. Users may use personal devices, together with the standard digital television device in order to watch the presentation, at the same time that he/she is able to capture a frame, and make annotations using, for instance, a pen-based device (Figure 1 (a)). The interaction activity is monitored by a set of techniques that stores locally specific data, such as characteristics of frames (histograms, color level, etc.), captured frames IDs and timestamps from scenes, coordinates of strokes, recognized handwritings, etc. All this information is classified into different layers, represented by the blocks composing the inverted pyramid in Figure 1 (b)-(g), which, in turn, represents the required knowledge about the content.

The whole set of techniques corresponds to a services composition schema (Figure 1 (h)), which is controlled by a metadata server, responsible to manage, collect and store in a representation model valuable information about the

interaction activity among user and content (Figure 1 (i)). One example of such composition is the combination of segmented scenes with the annotations made onto a specific frame; this association, specially, is done by the comparison of the timestamps present in all media.

After all metadata is available in the representation model, the engine module (Figure 1 (j)) processes the data, creating the means for personalization services (Figure 1 (k)). Therefore, when the user interacts with the content, the architecture is able to extract metadata from specific information that is generated by the techniques. The metadata, in turn, is used to inform content providers about user's preferences that can be later explored to support personalization services.

The multimedia content that is transmitted from the content providers, which can be in its original form or personalized, is also processed by an adaptation procedure (Figure 1 (l)). Firstly, this step considers the collection of context information from the environment, and later, the multimedia stream is adapted using a set of composite services managed by an adaptation server.

5. EXPECTED RESULTS

Among the expected results of this project, we can highlight the following:

1. Implementation of a prototype that adapts and personalizes multimedia content according to environment's restrictions and user's preferences.
2. Interaction capabilities between adapted/personalized content and users, using different devices.
3. Possibility to extend the system by adding new adaptation and/or personalization services.

6. ACKNOWLEDGMENTS

This work is being sponsored by UOL (www.uol.com.br), through its UOL Bolsa Pesquisa program, process number 20080129100700. The authors would like to thank the financial support from UOL and FAPESP.

7. REFERENCES

- [1] G. D. Abowd, M. Gauger, and A. Lachenmann. The Family Video Archive: an annotation and browsing environment for home movies. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 1–8, 2003.
- [2] L. Agnihotri, J. Kender, N. Dimitrova, and J. Zimmerman. Framework for personalized multimedia summarization. In *Proceedings of 7th. ACM SIGMM Int. Workshop Multimedia Information Retrieval*, pages 81–88, 2005.
- [3] M. Aurnhammer, L. Steels, and P. Hanappe. Integrating collaborative tagging and emergent semantics for image retrieval. In *Proceedings of WWW 2006 Collaborative Web Tagging Workshop (Online)*, 2006. Available at: <http://www.rawsugar.com/www2006/17.pdf>.
- [4] V. M. G. Barrios, F. Modritscher, and C. Gutl. Personalization versus Adaptation? A User-centred Model Approach and its Application. In *Proceedings of I-KNOW'05*, pages 120–127, 2005.
- [5] D. C. A. Bulterman. Animating Peer-Level Annotations Within Web-Based Multimedia. In *7th Eurographics Workshop on Multimedia*, pages 49–57, 2004.
- [6] D. C. A. Bulterman. Is it time for a moratorium on metadata? *IEEE Multimedia*, 11(4):10–17, 2004.
- [7] P. Cesar, D. C. A. Bulterman, and A. J. Jansen. An Architecture for End-User TV Content Enrichment. *Journal of Virtual Reality and Broadcasting*, 3(9), 2006.
- [8] T. Choudhury and S. Basu. Modeling conversational dynamics as a mixed-memory Markov process. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2004.
- [9] C. Dorai and S. Venkatesh. Computational media aesthetics: Finding meaning beautiful. *IEEE Multimedia*, 8(4):10–12, 2001.
- [10] N. Eagle. *Machine perception and learning of complex social systems*. PhD thesis, Massachusetts Inst. Technol., 2005.
- [11] J. Gemmell, A. Aris, and R. Lueder. Telling stories with my lifebits. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Amsterdam, The Netherlands, 2005.
- [12] W. Y. Lum and F. C. M. Lau. A Context-Aware Decision Engine for Context Adaptation. *IEEE Pervasive Computing*, 1(3):41–49, 2002.
- [13] M. G. Manzato, D. B. Coimbra, and R. Goularte. Multimedia content personalization based on peer-level annotation. In *Proceedings of the 7th. European Interactive TV Conference (EuroITV)*, 2009. (to appear).
- [14] M. G. Manzato and R. Goularte. Video News Classification for Automatic Content Personalization: A Genetic Algorithm Based Approach. In *Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 1–10, 2008.
- [15] M. G. Manzato, D. C. Junqueira, and R. Goularte. Interactive News Documents for Digital Television. In *Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 1–6, 2008.
- [16] M. G. Manzato, A. A. Macedo, and R. Goularte. Evaluation of Video News Classification Techniques for Automatic Content Personalization. *International Journal on Advanced Media and Communications (IJAMC)*, 3(4), 2009.
- [17] F. Nack and W. Putz. Designing annotation before it's needed. In *Proceedings of 9th. ACM International Conference on Multimedia*, pages 251–260, 2001.
- [18] S. N. Patel and G. D. Abowd. The ContextCam: Automated Point of Capture Video Annotation. *UbiComp 2004: Ubiquitous Computing, Lecture Notes in Computer Science*, 3205:301–318, 2004.
- [19] S. Venkatesh, B. Adams, D. Phung, C. Dorai, R. G. Farrell, L. Agnihotri, and N. Dimitrova. "You Tube and I Find" – Personalizing Multimedia Content Access. *Proceedings of the IEEE*, 96(4):697–711, 2008.
- [20] H. Zettl. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth, London, UK, 1999.

Audiovisual cultural heritage: Bridging the gap between digital archives and its users

Guido Ongena, Ph.D. student
Department of Media, Communication & Organisation
University of Twente, P.O. Box 217, 7500AE, Enschede
g.ongena@utwente.nl

ABSTRACT

This document describes a PhD research track on the disclosure of audiovisual digital archives. The domain of audiovisual material is introduced as well as a problem description is formulated. The main research objective is to investigate the gap between the different users and the digital archives. Next, design research is proposed as a methodology for this research. Lastly, the social and scientific relevance is briefly discussed.

Keywords

audiovisual heritage, user requirements, business model, design research

1. INTRODUCTION

Not long ago printed documents were seen as the most important historical assets of a country or nation. However, with the development of radio and television for more than 50 years, audiovisual (AV) content is becoming more and more part of the cultural heritage of a nation [1]. Hence, the awareness of the importance of this part of the nation's heritage is increased and is becoming more professional.

Once awareness is created, the next step is to preserve content to prevent it from damaging in the course of time and to safeguard it for future generations. The most common way to preserve such historical artifacts is to digitalize the content. The total amount of hours audiovisual recordings in Europe is estimated at 50 million of audio, video and film [2]. Most of it is in an analogue format. Main reason, next to preservation, to convert analogue material into a digital format is to fluent the exchange of material by adding additional information to this content [3]. In the process of digitization metatags are added to the content, which increases the searchability and retrieval of particular content. Moreover, digitalization of content can also operate as a driver to establish new services. The process of digitization is complex and gains new insights into properties like distribution, interoperability, integration with other collections, and metatagging (e.g. taxonomies) [4]. These are just few of the domains that are affected by the digitization.

For preservation reasons The Netherlands Institute for Sound and Vision was established in 1996. Main goal of this institute was to create an infrastructure for managing and the preservation of the audiovisual heritage of the Netherlands. This resulted, in 2006, in a large archive that contains 70% of the Dutch audiovisual heritage [5]. In 2009 the budget for the institute is estimated for around € 19 million [6], which increases yearly with a small amount. Now infrastructure is created, six organizations started a consortium called 'Images for the future'. Activities of this

consortium are to restore and preserve audiovisual material, digitization, copyrights management and disclosure of the content.

2. PROBLEM DESCRIPTION

Although countries differ among stages of digitalization of their audiovisual material, it is an important condition for disclosing the content. However, questions arise about the method to make the archive available for its user's potential. Governments and cultural organizations are discovering various opportunities for new ways to bring their material to the public. At the same time, new techniques are available to distribute audiovisual content to its users.

For the transmission of audiovisual content, a high capacity connection is another important condition. In The Netherlands 78% of the population has a broadband connection in 2008 [7]. Moreover, the internet has more opportunities for personalization of content and an increase of web based applications with an uplifting ease of use.

When both conditions are sufficient, pilots can be started around intelligent services or applications for users. In the domain of audiovisual material five actors or users are important to stipulate. It is important to distinguish various users, because of different kind of needs and demands of these applications. The five actors with respect to audiovisual cultural heritage are [4]:

- Broadcasters (Public/Commercial, native/foreign)
- Profit organizations (DVDs, CDs, internet, commercial screenings)
- Non-Profit organizations (museums, film festivals)
- Education (primary, secondary, higher)
- General Public

Although these user groups are clearly defined several questions arise when investigating these groups and their relationship with audiovisual content. What kind of content do they need? Hence what are the user requirements per group? What is the influence of context in this retrospect; what is the demand of different users in different contexts? Moreover, what is their behavior to the digital archives? Next to the user requirements, what addresses the demand; it is important to study the possibilities of technologies and the match with the user requirements. In addition, which business model can be made upon both perspectives (demand and possibilities).

3. RESEARCH OBJECTIVE

This PhD research will be conducted in The Netherlands and there will be an intensive collaboration with Belgium. The

proposal is funded by ICTRegie¹ and IBBT². A Belgium PhD candidate will conduct the same research, differing in that he will describe the situation in Flanders. Section 3.1 describes the granted proposal where the content has been specified by the two PhD candidates. Additionally, a more fundamental objective or perspective is addressed in section 3.2.

3.1 Information science research

The aim of this research is to give answers to the questions, which are described at the end of section 2. Thus, this PhD project investigates the user requirements with respect to audiovisual content. Moreover it tries to identify technological possibilities and their contribution to the (better) use of audiovisual material. From both approaches a business model will be created, which contains critical success factors specifically when disclosing audiovisual material. Figure 1 depicts the research project and the related actors and applications. The left side of this image represents the digital archive as such. This large database of audiovisual content can be used by services, which resembles second skin for delivering content to the user. This user is stated at the right side. As stated above different users are formulated. These users have needs, which are important to develop a service or application. Thus, the service has to match with the needs of the different users.

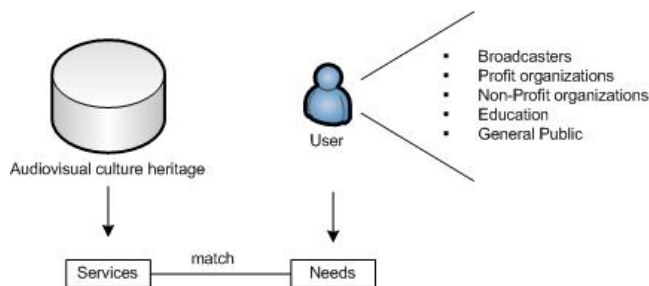


Figure 1. Positioning of the PhD research project

Much current research is done on the technical side of the digital archives of audiovisual material (e.g. digitization, indexing, video formats, automated tagging). The distinguishing part of this research lies in the fact that the next step has to be taken, to disclose the archives for its users. What will be accomplished with this research is to uncover the potential of the audiovisual heritage and to extend the discussion from a technological point of view to a more user perspective point of view with respect to audiovisual content. Moreover, different theories will be used from other domains. This expands the validation or discussion of these widespread theories (e.g. technology acceptance model [8]).

Although this research mainly studies the user perspective of audiovisual content, technological issues will be addressed. For instance, it can be interesting to investigate the integration of a folksonomy (metatagging by users) with a taxonomy (metatagging by professionals). Hence, technological issues will

¹ The goal of the *Netherlands ICT Research and Innovation Authority* is to develop a national ICT research and innovation strategy that is aimed at strengthening the ICT knowledge infrastructure of the Netherlands and maximising the benefits for society and the economy.

² IBBT (Interdisciplinary Institute for Broadband Technology) is an independent research institute founded by the Flemish government to stimulate ICT innovation.

be addressed when necessary. An important issue in this perspective is the use of AV material on different platforms. The use of mobile devices is emerging and inherent the use of internet applications on these devices. Therefore we distinguish the following platforms:

- Television (linear, video on demand)
- Internet (websites, portals)
- Mobile device (smartphone, netbook)
- Hardcopy (CD, DVD, Blu-ray)

3.2 Research paradigm

In the above described research a fundamental framework comes up, which lies underneath and has more a meta research character. The research revolves around the utility of a service. This idea of research through utility can be found in the information systems research framework. Behavioral science investigates theories for explaining or predicting human behavior. Design science on the other hand investigates the capabilities and boundaries of humans and organizations and tries to expand these by creating new and innovative applications or services. Hence, design research revolves around creating [9]. Figure 2 presents the design science framework as documented by Hevner et. al.

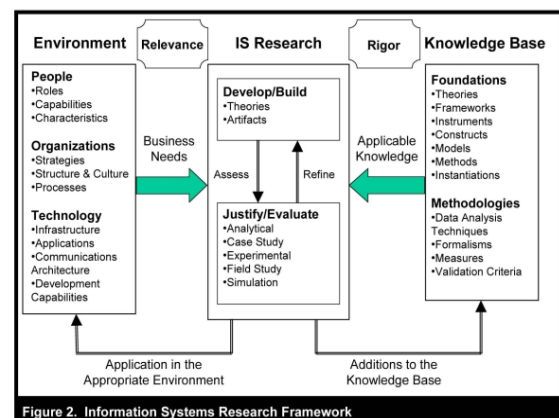


Figure 2. Information Systems Research Framework [9]

The main concept is that data is collected from the environment, which can be used for the development. As are also theories and frameworks from the knowledge base which in this case can be seen as the scientific world. Questions arise such as how the research objective is related to the above framework. As can be seen in Figure 1 this research revolves around the users of audiovisual material. The basic idea of design research will be used in all steps of the conducted research. Thus, when investigating the user needs theories and instruments of the knowledge base will be used as data from the environment where we live in.

However a more fundamental issue is created upon this framework. The doctorate will include one or more chapters about the assessment or evaluation of this framework. Thus, when executing the research as is proposed in section 3.1, remarks will be noted about this framework. This is done in order to improve or extend the theoretical framework.

Now the research objective and positioning is set, the methodology will be discussed in further detail below.

4. METHODOLOGY

This PhD research is currently at the very beginning, thus this part is still in an early stage. However, the method for collecting data is important, because it determines for a large part the internal and external validation of the research that is conducted. The overall method or research framework can be characterized as design research [9], which is explained in previous section. Below, the methods per stage will be briefly discussed.

4.1 User needs and technical possibilities

To gain full insight in the behavior of consumers with respect to the use of audiovisual content a survey method is proposed. The main reason for this method is for extrapolating the results to a broader society. In addition to the survey, datamining can be a suitable additional method. Much data is stored in log files. These log files contain the queries of users who searched for audiovisual content.

For acquiring knowledge about the state of the art techniques or applications that are available for bringing the AV-content to its users, an in depth literature study will primarily be conducted.

When both segments (user behavior and state-of-art techniques) are completed it is necessary to combine both perspectives to ensure one fits to the other. In order to match these perspectives we propose a vignette method. Respondents receive a short scenario, upon which they answer a couple of questions.

4.2 Services and viable business model

As the design research stipulates, the basic idea of the framework involves utility. Therefore, an application or service will be developed during this research project. Important factors in this are the user needs and the technological possibilities, which are formulated in section 4.1. Hence, these factors serve as input for a user based service.

Next to the development of a service, a business model is constructed from theory and data is gathered from user research. Users are asked in the scenario's about their thresholds and motives for using audiovisual content. This data will function as an input factor for a business model which cultural organizations can support for maintaining their businesses. For the development of a business model, the STOF model will be used as a framework [10]. This framework is validated in different sectors and gives a model for identifying critical success factors specifically for audiovisual content.

4.3 Justify and evaluate

After developing a service or application to one of the groups, an assessment is needed for refining the service. This will be done by means of experiments. These experiments involve the user experience and usability issues that are related to this service or application. The output of these experiments is used for the recommendation of the application.

5. SCIENTIFIC & SOCIAL RELEVANCE

Discussing the main contribution of this research to iTV and the justification why this contribution will lead to a PhD degree touches upon the scientific and social relevance (also in relation to iTV) of this research.

Although the scientific relevance is perceivable throughout the sections of the research objective and the methodology part (hopefully), it is hard to formulate a single abstract research question. This is caused by the methodology of design research, where environment and knowledge base touches upon each other within different stages of research. Nevertheless, this research can bring major knowledge to the academic society about the interaction between the audiovisual archives and its users.

In general, the social relevance of this research lies in the fact that culture organizations gain insights in the user needs, but also acquire knowledge about a viable business model. This is important because more and more organizations are shorted on government support and are reliable on commercial funds and customers. Hence, these organizations seek for better and viable business models.

The relation with iTV is unambiguous. This PhD research track is involved with numerous topics which are stated on the EuroITV website. To name a few topics that are relevant in relation to the research objective above: digital content production, web2.0, user-generated content, usability and user experience, business models, personalization, and user modelling.

6. REFERENCES

- [1] Bachimont, B., and Auffret, G. 1999. Audiovisual Cultural Heritage: From TV and Radio Archiving to Hypermedia Publishing. In Third European Conference and Research and Advanced Technology for Digital Libraries. BNF, Paris, 58-75.
- [2] Wright, R. 2008. Preservation of Digital Audiovisual Content. Consulted February 18, 2009: http://www.digitalpreservationeurope.eu/publications/briefs/audiovisual_v3.pdf
- [3] Caldera-Serrano, J. 2008. Changes in the management of information in audio-visual archives following digitization: Current and future outlook. *Journal of Librarianship and Information Science*. 40, 1 (March 2008), 13-20. DOI=<http://dx.doi.org/10.1177/0961000607086617>.
- [4] Oomen, J., Verwayen, H., Timmermans, N., Heijmans, L. 2008. Images for the Future: Unlocking Value of audiovisual heritage.
- [5] The Netherlands Institute of Sound and Vision. Consulted February 19, 2009: <http://instituut.beeldengeluid.nl/index.aspx?ChapterID=8532>
- [6] Ministry of Education, culture and science. 2009. Rijksbegroting 2008. [dutch]
- [7] Statistics Netherlands. 2008. ICT gebruik van personen naar persoonskenmerken. [dutch]
- [8] Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 319-339
- [9] Hevner, A.R., Salvatore, T.M., Park, J., Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly*. 28, 1 (March 2004), 75-105
- [10] Bouwman, H., DeVos, H., Haaker, T. 2008. Mobile Service Innovation and Business Models. Springer-Verlag, Berlin

The Next Generation of Multimedia Authoring Tools: Telling Stories and Commenting on Media

Rodrigo Laiola Guimarães
CWI: Centrum Wiskunde & Informatica
Science Park 123
1098 XG Amsterdam, The Netherlands
+31 20 592 93 33
rlaiola@cwi.nl

ABSTRACT

Next generation of multimedia authoring tools will be targeted to the end-users. This assumption is the starting point of my thesis work, which deals with the underlying technological infrastructure and the needed theoretical models for realizing next generation multimedia authoring tools. My argumentation is based in one common trend. While in the past the multimedia lifecycle was authoring-distribution-consumption, in the current landscape viewers have become active nodes in the networked media chain. Now, they have become, apart from consumers, producers and distributors. Thus, they need a new set of authoring and sharing tools to satisfy their needs. At this stage in my PhD, after a detailed review of related work and systems (e.g., YouTube) and the implementation of a number of prototypes (e.g., caption interface and pan & zoom authoring tool), I can identify the topic area of my thesis and the following next steps to achieve it.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems - *Human factors*. H.5.2 [Information Interfaces and Presentation]: User Interfaces - *User-centered design*.

General Terms

Design, Experimentation, Human Factors.

Keywords

User-centric media, End-user authoring systems.

1. INTRODUCTION

Traditionally, the production and distribution of digital TV content follow a one-way approach. First, high-quality material is captured using professional audiovisual devices. Then, the raw material is edited using tools, which operate at a low syntactic level, for example manipulating video as a sequence of frames and streams of un-interpreted audio. The TV programs are then authored aggregating the processed media assets into one

presentation, in which the layout characteristics of the visual elements are defined, and the interactivity points handled. Finally, the produced content is delivered through broadcasting stations to the end-users' device for consumption [1].

In contrast with this traditional media flow, contemporary research and the explosion of new services in the Web show an interesting trend, in which the viewer is not the end of the content value chain anymore. Instead, the viewer is considered to be an active node in the production-distribution-consumption chain. More recently, triggered by the widespread availability of digital recording devices as well as display and rendering devices, the viewer has become a potential content producer and content distributor. User-generated content websites (e.g. YouTube¹), web-based authoring tools, such as JumpCut² and the work reported by Shamma [8], and viewer-side enrichment solutions [3][4] are examples of this trend. Table 1 presents a comparison between the traditional and the user-centric media flows.

Table 1. Media flow landscape.

	Traditional	User-centric
Content Authoring	Professional mindset	Entertainment-oriented
	Authoring tools prioritizes technical aspects	Authoring tools prioritizes authors' needs
Content Delivery	Client-Server approach	Hybrid approach (e.g. Client-Server and P2P)
	one-to-many	one-to-one
		one-to-some
		one-to-many
Content Consumption	Passive Viewers	Proactive Viewers
	Restricted interaction (e.g. turn on/off, channel switch and playback commands)	Rich interaction (e.g. chat and voice channels, content enrichment)

The final goal of my thesis is to determine how technology can empower users to become active nodes in this new networked media lifecycle. This paper is organized as follows. Section 2

¹ <http://www.youtube.com>

² <http://www.jumpcut.com/>

discusses about the problem that the proposed research addresses, as well as the main objectives of this work. Section 3 presents the methodology I am following to achieve the objectives, indicating the current stage of my thesis work. Then, Section 4 is dedicated to show some end-user oriented prototypes I have been working in during the last year. Finally, Section 5 reports on current findings and the future work in the continuation of my PhD.

2. OBJECTIVE OF THE THESIS

Authoring systems for interactive digital television, and for multimedia content in general, assume the author to be seated in front of a workstation on the broadcaster side and with a professional mindset. These systems have tackled the authoring complexity problem in different ways. While some commercial systems artificially reduce the authoring complexity by limiting the available features (e.g. JAME Author³, Cardinal Studio⁴ and AltiComposer⁵), several research systems have attempted to provide support for a wider range of authoring needs [2][6]. Apart from functionality limitations, systems for interactive digital TV are generally technology-centered, that is, the design of standards, infrastructures and authoring tools normally prioritizes the technical aspects over the real authors' needs (see Table 1). Therefore, authors have to spend a reasonable amount of time in "recycling" whenever a new technological feature is incorporated.

The establishment of Web technologies had major implications on content creation paradigms and methods: from the traditional professional creation of content to a more fresh, immediate, and incidental creation of user-generated content (UGC). Web-based authoring tools generally require a less professional mindset than traditional authoring systems and they tend to provide the functionality the user needs one click away. In addition, end-user enrichment systems, such as YouTube Annotations, the Ambulant Annotator [4] and the work reported for the Brazilian Digital TV System (SBTVD-T) [3], are capturing the attention of end-users.

Being simplistic, the content authoring-distribution-consumption chain can be seen as in Figure 1.

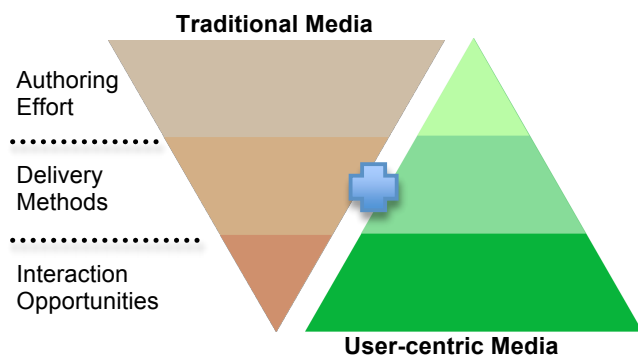


Figure 1. Comparison between Traditional and User-centric Media.

The first layer in the figure represents the media authoring effort. Traditional authoring tools provide a wide range of possibilities,

³ <http://jame.tv/>

⁴ <http://www.cardinal.fi>

⁵ <http://www.alticast.com>

but they usually require training, a steep learning curve and a big effort from the author – represented as a bigger area in the traditional media landscape. On the other hand, authoring systems intended for end-users (e.g. enrichers and annotators) tend to be easy to use (less authoring effort, represented by a smaller area in the user-centric side), but they usually offer a limited range of authoring capabilities.

Talking about media delivery methods and sharing capabilities (middle layer in Figure 1), while the traditional approach was to distribute media using a client-server or a broadcast infrastructure, the success of Web-based television portals and of Peer-to-Peer (P2P) technologies have demonstrated that end-users have changed the way in which they find and consume audiovisual media. In terms of delivery aspects I would not claim that the variety of delivery methods in the user-centric landscape is bigger or smaller than in the traditional landscape. However, we can assure that the new media landscape by combining both approaches offers a variety of delivery and sharing mechanisms broader than ever before.

Finally, the interactive opportunities of the television viewers have been increased in the last years. Contrary to the traditional role of television viewers as passive spectators, with limited interaction (represented by a small area in the traditional media landscape), there are strong indications that there is room for more active television watching. Therefore, we can say that the user-centric paradigm added a wider variety of interaction opportunities to the new media landscape (e.g., annotate, enrich, and share television content while watching). It is important to note, though, that according to observers [7][5], it would not be the flavor of interactivity with complex applications that require concentration, long time spans, and distraction from television viewing. Instead, new interactivity models particularly designed for television are needed.

Still, as reported by Jensen [7], in the new media landscape the relations between media and consumers are not sided. That is, we do not have passive viewers consuming only mainstream media and active viewers only producing active media. Every consumer is potentially an active creator and a passive consumer and he is able to switch between the two roles.

It is important to state that contrary to how the metaphors in Figure 1 look like, the media flows are not in opposition to each other. Actually, we can consider the new media landscape as the sum of both media flows. Therefore, in this thesis I do not claim that one of the two paradigms is better or worse than the other, or that even one or the other will disappear in the coming future. The point is that media has swiftly evolved in the past years in a way never seen before, hence people's experience when interacting with this new media. And this change is mostly due to the user-centric media paradigm.

The main objective of my thesis is to understand, develop, and model the underlying technology that helps viewers to play an active role in the media production-distribution-consumption chain.

3. METHODOLOGY

In order to achieve the objectives, I have been following the workflow shown in Figure 2. I started with a theoretical and a practical groundwork in order to better understand the underlying problems. On the one hand, I investigated existing authoring services intended for end-users. On the other hand, I reviewed and

analyzed related research in the fields of multimedia authoring systems. This groundwork resulted in the development of a number of prototypes (c.f. next section) that helped me in finding new authoring models, in which the end-user plays a major role. This is the current stage of my thesis.

The next steps consist on a continually collaboration with human-centered interactive television groups⁶ for identifying authoring models based on the potential needs of representative actors. Then, I hope to identify actual services and specify the underlying technology for the next generation authoring systems. I will do this in the context of the European Project Ta2⁷. The intention is to validate each of these steps from a user experience evaluation compared to other systems, and from the business perspective. Based on the findings I will hopefully have the necessary material to conclude my PhD thesis.

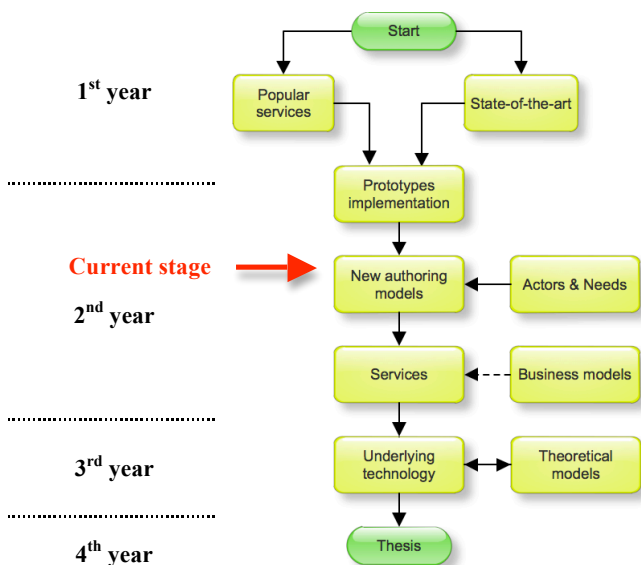


Figure 2. Thesis workflow.

4. SELECTED EXAMPLES

During the prototypes implementation phase I have been working in a set of authoring systems focused on end-users. For better illustrate the scenarios in which these systems are applied consider the Sneijders: the couple Anthonius (46) and Seletje (44), and their two children, a girl named Rebekka (13), and a boy called Rykaard (11). This family has recently moved from one end of the country to the other, the children's grandparents, Peter (77) and Alita (73) are not able to come to visit them very often.

4.1 The Pan/Zoom Effect

Consider the photo depicted in Figure 3, taking when Grandma Alita was younger. She might want to tell a story about her old and good times, and share it with her non-collocated grandchildren. Life could be given to this still photograph by zooming in/out on each member (subject of interest) while she explains who is who in the photo, and panning (moving the focus) from one to another member. Although grandma and

grandchildren are separated spatially, the visual dynamics adds sentiments to the experience, in comparison to traditional means to stay in touch, such as a telephone call.

Apart from encouraging elderly people to produce and augment their stories – either to help them to document their historic memories and experiences in general – in this scenario I highlight the importance of providing users a way to define a dynamic storyline composed by audio commentaries and visual effects over images, giving the feeling of motion, and keeping the viewers visually entertained.

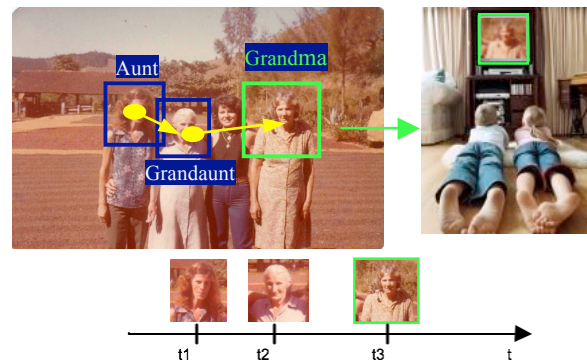


Figure 3. Storytelling a family photo.

4.2 MyVideos

The band at the local junior high school recently gave its Spring concert. The band consisted of a collection of amateur musicians and Rebekka was one of them. At the concert, many of the families brought both still and video cameras. The school also had an A/V setup that made a continuous recording of the concert with high-quality audio.

The purpose of each of the video collections is primarily personal: each family is interested in capturing their own child, but also enough context information from the concert to provide some background. Not everybody films all the time. Each family is interested in creating a video fragment for the personal family archives, or short clips that can be sent to family members who were not able to be at the show. For example, grandma Alita would like to see what Rebekka is up to with the flute.

All of the parents are used to multi-camera concert videos: they would like to make use of the material of others when it is appropriate – either to show the shot they missed or to provide a more compelling video by having multiple views. Since, however, this is a musical event, the audio transitions should be minimal – perhaps being driven by the master soundtrack recorded by the school. In addition to the individual collections of content, the school may also be interested in creating an archival of the video, in which each participant is shown, as well as some background shots of the band and the crowd.

The application enables people to create videos of the concert based on the raw footage of any contributor happy to post their video to a common site. The site application supports the subsequent organization and tagging of the video, and the user production of their films. Thus helping Seletje and Anthonius to create their own story around Rebekka's concert.

4.3 Captioner

Consider that granddad Peter and grandma Alita received a personalized video with the concert of their granddaughter

⁶ University of Salzburg, CUO Leuven and the University College of London.

⁷ <http://www.ta2-project.eu/>

Rebekka. After a couple of minutes watching the presentation, granddad Peter, who suffers from hearing loss, got frustrated and decided to stop watching the video. Taking notice of this situation, his son Anthonius decides to make use of the captioning system shown in Figure 4.

This captioner system allows anyone to add captions and subtitles to arbitrary videos. The user can select a video and then, using the input area show in the figure, add the captions. The result can be shared with people near and far. The system provides many useful features including: keyboard shortcuts, subtitles translation support, styling and a tight temporal coupling with the video media in the presentation.

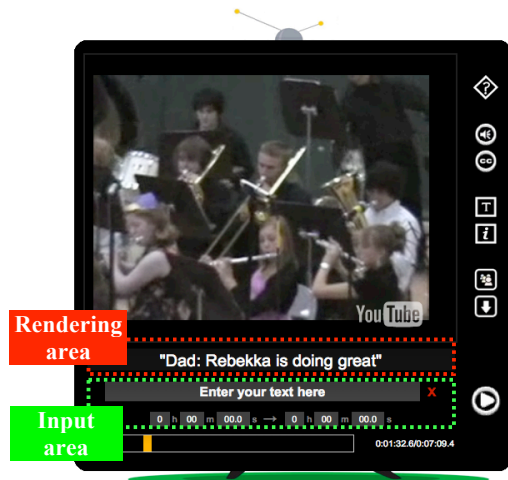


Figure 4. Captioning a video.

4.4 Annotator

During a Spanish movie, Anthonius decides to share relevant parts of the video with his parents Peter and Alita. Using a personal device (e.g., iPhone), Anthonius can quickly annotate and enrich a piece of the video [4], which then can be shared with his parents. The enrichments over the video are done while watching the content. Figure 5 illustrates this scenario.



Figure 5. Sharing the television-watching experience [4].

The authoring in this case consists of overlaying a personal navigation structure highlighting portions of interest, and adding a number of voice-over annotations to some of the fragments — such as ‘Do you remember this movie?’ or ‘We have been in that place before’. He then sends a recommendation message to his parents with a pointer to “his” version of the video. Note that such enrichments do not generate a new version of the video; they create a (set of) content wrapper(s) containing a link to the base video, along with a navigation map and a set of — possible personalized — annotations.

5. FINAL REMARKS

From the scenarios presented in the previous section it is possible to derive two clusters of authoring systems targeted to end-users:

the ones intended for creating personal stories and the ones intended for the creation of personal media. The characteristics of these two paradigms are shown in Table 2.

Table 2. Report on end-user authoring systems.

	Personalized Story	Personalized Media
Type of content	Normally UCG	Normally existing media
Actions	Determine the media navigation path	Add/remove/modify media items (overlays)
Objective	Create a story with a narrative	Comment/give personal view
When	After capture	During consumption
Example	Pan/Zoom & MyVideos	Captioner & Annotator

This categorization represents the first results in my thesis work, as a model for next generation multimedia authoring systems. Based on these findings, the next steps of my PhD consist on the design, development, and testing of new services, as indicated in the methodology presented above.

6. ACKNOWLEDGMENTS

This work was supported by EU FP7-ICT project TA2. My thanks to Pablo Cesar and Dick Bulterman for their useful comments.

7. REFERENCES

- [1] Bulterman, D. C. A. 2007. User-centered control within multimedia presentations. In *ACM/Spinger Multimedia Systems Journal*, Volume 12, Numbers 4-5, pp. 423-438.
- [2] Bulterman, D. C. A. and Hardman, L. 2005. Structured multimedia authoring. In *ACM TOMCCAP*. 1(1): 89-109.
- [3] Cattelan, R. G., Teixeira, C., Goularte, R. and Pimentel, M. G. C. 2008. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. In *ACM TOMCCAP*, Volume: 4 Issue: 4, November 2008.
- [4] Cesar, P., Bulterman, D. C. A., Geerts, D., Jansen, A. J., Knoche, H. and Seager, W. 2008. Enhancing Social Sharing of Videos: Fragment, Annotate, Enrich, and Share. In *Proceedings of the ACM MM 2008*, pp. 11-20.
- [5] Chorianopoulos, K. 2008. User Interface Design Principles for Interactive Television Applications. In *International Journal of Human-Computer Interaction*, 24(6), pp. 556-573.
- [6] Guimarães, R. L., Costa, R. M. R., and Soares, L. F. G. 2008. Composer: Authoring Tool for iTV Programs. In *Proceedings of EuroITV 2008*, pp. 61-71.
- [7] Jensen, J. 2008. The concept of interactivity -- revisited: four new typologies for a new media landscape. In *Proceedings of UXTV 2008*, pp. 129-132.
- [8] Shamma, D. A., Shaw, R., Shafton, P. L., and Liu, Y. 2007. Watch What I Watch: Using Community Activity to Understand Content, In *Proceedings of the International Workshop on Multimedia Information Retrieval*, pp. 275-284.

LIST OF CONTRIBUTORS

POSTERS

Aroyo Lora, Vrije Universiteit Amsterdam
Bressan Graça, Polytechnic School of University of São Paulo
Concolato Cyril, Télécom Paris Tech
Conconi Alex, TXT Polymedia
de Fez Ismael, Polytechnic University of Valencia
Dietze Stefan, Open University
EinHui Kim, Department of Information and Communication, KAIST
Förster Florian, ICT&S Center, University of Salzburg
Fraile Francisco, Polytechnic University of Valencia
Gambi Ennio, Università Politecnica delle Marche
Guerri Juan Carlos, Polytechnic University of Valencia
Hofstätter Jörg, ovos realtime 3D GmbH
Hurtado-Dueñas José Antonio, Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones
Igelsböck Judith, ICT&S Center, University of Salzburg
Kaptein Annelies, Stoneroos Interactive Television
Korhonen Ahti, University of Jyväskylä
Kranzer Jochen, ovos realtime 3D GmbH
Lobato Vanesa, Fundación CTIC
López Gloria, Fundación CTIC
Mäkelä Tiina, VTT Technology Centre of Finland
Münchau Björn, meta-morph.tv
Munchurl Kim, Department of Electrical Engineering, KAIST
Murayama Yuko, Iwate Prefectural University
Nixon Lyndon, STI International
Nufer Christoph, Institut für Rundfunktechnik
Obrist Marianna, ICT&S Center, University of Salzburg
Oksman Virpi, VTT Technology Centre of Finland
Pacola Alves Luiz Gustavo, Polytechnic School of University of São Paulo
Padilla-Conejo Rocío, Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones (CITIC)
Palmisano Davide, Asemantics
Peláez Víctor M., Fundación CTIC
Pinto Maia Luis Miguel, Novabase Digital TV
Rascioni Giorgio, Università Politecnica delle Marche
Ruiz-Villalobos Bernardo, Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones
Saito Yoshia, Iwate Prefectural University
Salinas J. R., Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones
Santos da Silva Fábio, Polytechnic School of University of São Paulo
Shinjee Pyo, Department of Information and Communication, KAIST
Spinsante Susanna, Università Politecnica delle Marche
Strzebkowski Robert, Beuth University of Technology Berlin
Sturm Petra, ovos realtime 3D GmbH
Tammela Antti, VTT Technology Centre of Finland
Tscheligi Manfred, ICT&S Center, University of Salzburg
Vignaroli Luca, RAI Radiotelevisione Italiana
Yankova Milena, Ontotext Lab, Sirma Group Corp

TUTORIALS

Alliez Damien, NDS Technologies France
Friedrich Oliver, Fraunhofer Institute FOKUS
Geerts David, CUO/IBBT, K.U.Leuven
Knoche Hendrik, University College London
Lugmayr Artur, Tampere University of Technology
Obrist Marianna, University of Salzburg
Seeliger Robert, Fraunhofer Institute FOKUS
Swings Koen, Zappware

WORKSHOPS

Bhachu Amritpal S, University of Dundee
Bleumers Lizzy, IBBT-SMIT VUB
Cesar Pablo, CWI
de Sousa Fernandes Marcelo, Federal University of Paraiba
de Souza Filho Guido Lemos, Federal University of Paraiba
Epelde Gorka, VICOMTech – IK4
Förster Florian, University of Salzburg
Freeman Jonathan, University of London
Godoy Piccolo Lara Schibelsky, CPqD – Telecommunications Research and Development Center
Harboe Gunnar, Motorola
Hassenzahl Marc, Folkwang University, Campus University of Duisburg-Essen
Hesselman Cristian, Novay
Igelsböck Judith, University of Salzburg
Izquierdo Ebroul, Queen Mary, University of London
Jacobs An, IBBT-SMIT VUB
Knoche Hendrik, University College London
Kommonen Kari-Hans, Media Lab, University of Art and Design Helsinki
Lessiter Jane, University of London
Miotto Andrea, University of London
Nack Frank, HCS – University of Amsterdam
Obrist Marianna, University of Salzburg
Palmeira André Felipe, Federal University of Paraiba
Passino Giuseppe, Queen Mary, University of London
Pataca Moutinho Daniel, CPqD – Telecommunications Research and Development Center
Patras Ioannis, Queen Mary, University of London
Pearmain Alan, Queen Mary University of London
Rebaque Pablo, Barcelona Media – Center of Innovation
Robinson Peter, University of Cambridge
Savrami Katia, University of Patras
Shepherd Keir, HP Labs
Springett Mark V., Middlesex University
Timoteo de Souza Dênio Mariz, Federal University of Paraiba
Tuomi Pauliina, Tampere University of Technology
Ursu Marian F, University of London
Wiklund-Engblom Annika, iDTV Lab, MediaCity, Åbo Akademi University
Williams Doug, Broadband Applications Research
Yannopoulos Angelos, National Technical University of Athens

DEMOS

Antonini Roberto, Telecom Italia Lab
Bellekens Pieter, Eindhoven University of Technology
Berrani Sid-Ahmed, France Telecom

Flórez Julián, VICOMTech
Geymonat Marina, Telecom Italia Lab
Kaptein Annelies, Stoneroos Digital Television
Manson Gaël, France Telecom
Marrow Paul, BT Innovate
Montanari Roberto, University of Modena-Reggio Emilia
Muenchau Bjoern, meta.morph GbR
Naturel Xavier, France Telecom
Neumann Jonas, meta.morph GbR
Newell Chris, BBC Research & Development
Olaizola Igor G., VICOMTech
Oyarzun David, VICOMTech
Perrero Monica, University of Torino
Simeoni Rossana, Telecom Italia Lab
Strzebkowski Robert, Beuth Hochschule für Technik
Ugarte Alex, VICOMTech
Van Kerckhove Geert, Stoneroos Digital Television
Voß Alexander, Microsoft EU Innovation Center

ITV IN INDUSTRY

Bächle Mirja, facit digital
Damian Kathrin, facit digital
De Bruecker Ward, KOC – Kenniscentrum Hulpmiddelen
Gielen Frank, Siruna
Jorge Hernández Pablo, Telefónica I+D
Lai Fon-Lin, Institute for Infocomm Research
Lapointe Pierre-Alexandre, YU CENTRIK
León Oscar, Optiva Media
Meyers Paul, Paratel
Navarro-Prieto Raquel, Barcelona Media Innovation Centre
Rebaque Pablo, Barcelona Media Innovation Centre
Van den Breede Gerrit, KOC – Kenniscentrum Hulpmiddelen
Van Deursen Davy, Siruna
Vidarthi Jay, YU CENTRIK
Yau Wei-Yun, Institute for Infocomm Research

DOCTORAL CONSORTIUM

de Resende Costa Romualdo Monteiro, Departamento de Informática – PUC-Rio
Ferreira Moreno Marcio, Departamento de Informática – PUC-Rio
Joly Ana Vitória, University of Brighton
Laiola Guimarães Rodrigo, CWI: Centrum Wiskunde & Informatica
Manzato Marcelo G., University of Sao Paulo
Ongena Guido, University of Twente